TITLE

From paleness to albinism: Contribution of *OCA2* exon 10 skipping to hypopigmentation

Short Title: *OCA2* exon10 skipping and pigmentation

Elina Mercier[1], David-Alexandre Trégouët[2], Sébastien Campagne[3], Vincent Michaud[1,4], David J. Green[5], Shahram Mesdaghi[6,7], Orane Le Gallais[1], Orphée Jouannigot-Castano[1], Daniel J. Rigden[6], Catherine Estay-Ahumada[8], David Hicks[8], Panagiotis I. Sergouniotis[5], Benoit Arveiler[1,4], Sophie Javerzat[1*]

1 Rare Diseases Genetics and Metabolism, INSERM U1211, University of Bordeaux, Bordeaux, France

2 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, Bordeaux, France

3 Univ. Bordeaux, CNRS, INSERM, ARNA, UMR 5320, U1212, Bordeaux, France

4 Molecular Genetics Laboratory, Bordeaux University Hospital, Bordeaux, France

5 Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom,

6 Institute of Systems, Molecular & Integrative Biology, Biosciences Building, Crown Street, Liverpool, United Kingdom

7 Computational Biology Facility, MerseyBio, University of Liverpool, Crown Street, Liverpool, United Kingdom

8 Institute for Cellular and Integrative Neuroscience, CNRS UPR 3212, Strasbourg, France

* Corresponding author

E-mail: sophie.javerzat@u-bordeaux.fr https://orcid.org/0000-0003-3497-0542

Abstract

Pathogenic variants in the *OCA2* gene result in oculocutaneous albinism. In humans and several other mammalian species, *OCA2* is transcribed into two coding mRNAs, a major transcript that encodes the full-length protein and a minor transcript that skips in-frame exon 10 and has no identified function. By contrast, rodents and non-mammalian vertebrates only express the major transcript. Several rare variants of human *OCA2* have been reported to be pathogenic due to increased exon 10 skipping. To understand the vulnerability of human *OCA2* to exon 10 skipping and its impact on pigmentation, we first functionally tested a series of rare missense variants spread across exon 10. We found that each variant significantly influences the skipping ratio either positively or negatively. By combining human and murine sequences in functional assays, we found that the skipping ratio depends on particular exonic and intronic combinations. Next, we focused on the most frequent exonic single nucleotide variant of human *OCA2*, rs1800404-T (c.1065G>A/p.Ala355=), located in exon 10. We show that this variant significantly promotes exon 10 skipping on its own and exhibits an additive effect in *cis* to previously identified splicing variants likely contributing to their pathogenicity. Association studies reveal that rs1800404-T is tightly associated with lighter skin and hair pigmentation in a representative European population, as has been observed in other populations around the world. Our structural models of the skipped transcript-derived protein support causality by predicting that it could exert a dominant-negative effect, which is consistent with a dose-dependent hypopigmentation response in both pathological and physiological contexts. Overall, the modulation of *OCA2* exon 10 skipping by both benign and pathogenic variants provides a basis for improving the genetic diagnosis of albinism, and paves the way for research into the molecular and evolutionary mechanisms behind human pigmentation diversity.

48 # Introduction

49

50 Like most genetic conditions, albinism displays a wide range of phenotypic manifestations, which can
51 make diagnosis challenging. Clinical diagnosis predominantly relies on quantitative assessment of
52 specific ophthalmological abnormalities as well as skin and hair hypopigmentation - although the latter
53 may go unnoticed (1,2). Pathogenic variants that cause albinism express themselves as a continuous
54 spectrum from extreme whiteness and severe visual impairment to marginally subnormal vision and
55 mild sensitivity of the skin to UV-light. Notably, most genes involved in non-syndromic oculocutaneous
56 albinism (OCA) are also associated with fine-tuned modulation of pigmentation in healthy individuals.
57 This results in single nucleotide variants (SNVs) in these genes being exploited in forensic DNA profiling
58 to predict skin, hair and eye colour (3,4). Of note, a quarter of the SNVs selected in the last version of
59 the multiplex forensic phenotyping toolkit HirisPlex-S (10 out of 41), are in *OCA2* or in its enhancer
60 sequence located in *HERC2* (5), illustrating a major role of *OCA2* in modulating pigmentation.

61 *OCA2* encodes the OCA2 protein (previously called P protein), a melanosomal transmembrane anion
62 channel that is instrumental in regulating pH within the pigment synthesizing organelle (6).
63 Importantly, its contribution to melanogenesis appears equivalent in the retinal pigment epithelium
64 (RPE) and in cutaneous melanocytes (7). Lack of OCA2 protein markedly impairs melanogenesis
65 mediated by the three tyrosinase-related enzymes, Tyrosinase, Tyrosinase-related protein-1 and
66 Dopachrome tautomerase (Tyrosinase-related protein-2), respectively encoded by *TYR*, *TYRP1* and
67 *DCT/TYRP2*. Of note, the OCA2 protein has been reported to have a role in the processing and
68 trafficking of Tyrosinase from the endoplasmic reticulum to the melanosome (8). Loss-of-function
69 mutations in *OCA2* thus have a direct epistatic effect on the Tyrosinase encoding gene *TYR*. Once in
70 the melanosome, providing the pH is optimally controlled by several channels including the OCA2
71 protein, Tyrosinase catalyzes the initial and rate-limiting steps of the melanogenic biosynthetic
72 pathway by converting tyrosine into L-Dopa, a major factor implicated in retinal development. L-Dopa,

73 is abundantly produced in the embryonic RPE at the onset of melanogenesis and triggers critical steps
74 in early retinal development (9,10). Hence, individuals with *TYR*- and *OCA2*-related albinism exhibit
75 comparable levels of visual impairment and overlapping skin and hair pigmentation phenotypes (1).

76 Pathogenic variants in *OCA2* account for ~30% of genetic diagnosis of patients with clinically confirmed
77 albinism (11). The major transcript of *OCA2* carries all 24 exons. An alternative transcript that lacks the
78 in-frame exon 10 is expressed at low levels in healthy individuals. However, to date, the corresponding
79 putative protein remains undescribed. In a previous study, we showed that several rare variants
80 including synonymous variants of exon 10, significantly increase exon 10 skipping resulting in albinism
81 (12). Additional missense variants have been reported in exon 10 and are classified as variants of
82 unknown significance (VUS) due to poor *in-silico* prediction of protein misfolding. Their potential effect
83 on exon 10 skipping has not been investigated although splicing alterations are frequently triggered by
84 variants that lie in so-called vulnerable exons (13). These effects are however difficult to predict *in*
85 *silico*, reflecting the complexity of splicing *cis*-regulation (14). OCA genes do not escape the rule, with
86 an increasing number of splice variants being identified by functional assays that help resolve the
87 genetic diagnosis of patients (15,16).

88 Interestingly, basal levels of *OCA2* exon 10 skipped transcripts are only seen in some mammalian
89 species such as primates, but not in rodents. This contrasts with the structure/function similarity of
90 the orthologous OCA2 protein in human and mouse, that has made mouse mutant lines and cell-
91 derived models overexpressing the human OCA2 protein so useful to dissect its roles in melanogenesis
92 (6,17). It also questions a possible role of the skipped transcript in primates and other mammals that
93 express basal levels of this alternative mRNA.

94 Here, we investigated rare missense VUS that are distributed all along *OCA2* exon 10 and evaluated
95 their effect on splicing. We then asked what makes human *OCA2* highly vulnerable to exon 10 skipping.
96 Using different combinations of mouse and human constructs, we investigated the sequences that are
97 involved in the control of exon retention/skipping. We next focused on the most frequent exonic

variant of *OCA2* c.1065G>A/p.Ala355= (rs1800404-T) that lies in exon 10 and has previously been described as associated with increased levels of skipped transcripts (18) and with lighter pigmentation in Africans and other populations with dark skin (18–22). Here, we show that this variant enhances exon 10 skipping in a minigene assay and demonstrate its robust association with skin and hair paleness in the European population. These data are of interest not only in the context of genetic characterization of albinism and diagnosis, but also as a starting point for understanding the molecular and evolutionary mechanisms behind the physiological diversity of pigmentation in humans.

Results

Rare missense variants of *OCA2* exon 10 affect skipping profile

Six missense VUS of exon 10 (Table 1) that had been reported in different studies or identified in our OCA patient database were tested by minigene assay. Variant c.1048G>A/p.Val350Met has been reported twice in individuals with albinism (23,24) and, at the time of writing, is classified VUS in ClinVar. In our local database, this variant was found once in a Cypriot patient presenting with albinism who carried a deep intronic *OCA2* variant in *trans*. Variant c.1061T>C/p.Leu354Pro appeared only once in ClinVar as VUS. Variant c.1064C>T/p.Ala355Val has been reported once in a patient with albinism (25). It has been classified VUS twice and likely pathogenic once in ClinVar. Variant c.1081C>G/p.Leu361Val has been identified in one patient of our cohort who carried a deep intronic *OCA2* variant in *trans*. Variant c.1109T>C/p.Ile370Thr has been reported in two patients with albinism (23) and recorded as of Uncertain Significance, Benign or Likely Benign in ClinVar. Variant c.1114G>A/p.Asp372Asn is reported once in ClinVar as a VUS.

The positions of these variants are indicated in Fig 1a together with those of the nine variants functionally tested in our previous study (12). Allelic frequency, *in silico* predictions of splicing dysregulation or protein conformational changes as well as ClinVar reports are provided in Table 1.

OCA2 variants		In-silico predictions					GnomAD allele frequency (number HTZ and HMZ)		
c.	p.	SPiP	SpliceAI	Pangolin	Polyphen 2	Alphamissense	v3	v4 (non-UKB)	ClinVar report and variation ID
1048G>A	Val350Met	09.63 % (no effect)	0.02 (below threshold)	0.05 (below threshold)	0.984 (supporting pathogenic)	0.3939 (indeterminate)	9.2E-05 (14 HTZ-0 HMZ)	1.5E-04 (115 HTZ-2 HMZ)	VUS 1424156
1061T>C	Leu354Pro	09.76 % (no effect)	0.01 (below threshold)	0.01 (below threshold)	0.997 (supporting pathogenic)	0.8837 (supporting pathogenic)	6.6E-06 (1 HTZ-0 HMZ)	2.6E-06 (2 HTZ-0 HMZ)	VUS 2481092
1064C>T	Ala355Val	30.67 % (ESR element alteration)	0.01 (below threshold)	0.04 (below threshold)	0.981 (supporting pathogenic)	0.602 (indeterminate)	5.9E-05 (9 HTZ-0 HMZ)	2.3E-04 (181 HTZ-4 HMZ)	VUS 617804
1081C>G	Leu361Val	03.59 % (no effect)	0.01 (below threshold)	0.01 (below threshold)	0.9 (indeterminate)	0.178 (indeterminate)	6.6E-06 (1 HTZ-0 HMZ)	1.3E-05 (10 HTZ-0 HMZ)	Absent
1109T>C	Ile370Thr	05.05 % (no effect)	0.00 (no effect)	0.02 (below threshold)	0.038 (supporting benign)	0.0781 (moderate benign)	3.3E-03 (508 HTZ-4 HMZ)	9.7E-04 (755 HTZ-4 HMZ)	VUS Benign Likely benign 419971
1114G>A	Asp372Asn	09.3 % (no effect)	0.00 (no effect)	0.04 (below threshold)	0.086 (supporting benign)	0.0996 (moderate benign)	2.6E-05 (4HTZ-0 HMZ)	2.0E-05 (16 HTZ-1 HMZ)	VUS 2168452

Table 1: In silico predictions and allelic frequencies of missense variants included in this study. Nomenclature based on NM_000275.3. Predictions and frequency gathered from gnomAD: https://gnomad.broadinstitute.org/, SpliceAI, Pangolin, Alphamissense: https://spliceailookup.broadinstitute.org, mobiDetails, SPiP: https://mobidetails.iurc.montp.inserm.fr/MD/, ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/.

Abbreviations: HMZ, homozygous; HTZ, heterozygous; UKB, UK Biobank.

126 Overall, only c.1064C>T/p.Ala355Val displayed a significant score of potential alteration of a splicing
127 regulatory element. As for protein prediction*, in silico* analysis supported likely pathogenicity for
128 c.1061T>C/p.Leu354Pro only.

129 Results of the minigene assay are presented in Fig 1b-c. Of note, the common benign variant
130 c.1065G>A/p.Ala355= (highlighted in Fig 1a) was included in all constructs because it was in *cis* to the
131 tested missense variant in the corresponding patients from our cohort. In order to optimize
132 quantification of the skipping ratio, these assays were repeated 4 times in independent experiments
133 and the RT-PCR products run on an automated electrophoresis device coupled to data analysis
134 software. Details of the steps in data processing are provided in Fig S1. As illustrated, each of the six
135 missense variants showed a significant effect on splicing compared to the wild-type sequence taken as
136 a control (17% of skipping). Only variant c.1061T>C/p.Leu354Pro resulted in better retention of exon
137 10 compared to control. All other missense variants significantly increased the percentage of skipping
138 but to different extents. In order to further interpret these levels, they were compared to the profile
139 of synonymous variant c.1080C>T/p.Ser360= (65% skipping), previously shown pathogenic due to
140 excessive skipping (12). Among the newly tested missense variants, only c.1081C>G/p.Leu361Val
141 exceeded this threshold with 87% of skipped products. Variant c.1048G>A/p.Val350Met reached 53%
142 of skipping which is likely to contribute to its pathogenicity. The three remaining missense variants
143 showed weaker levels of enhancement reaching around 28% of skipping each. Of note,
144 c.1114G>A/p.Asp372Asn, despite lying 3 nucleotides before the exon/intron junction had the weakest
145 impact on loss of retention. All in all, the optimized minigene assay strongly supports pathogenicity
146 primarily due to splicing defect for c.1081C>G/p.Leu361Val. On the other hand, variant
147 c.1061T>C/p.Leu354Pro shows less skipping than the reference sequence. Its pathogenicity is
148 therefore likely to be the consequence of deleterious protein conformation changes as suggested by
149 the *in silico* analysis (Table 1).

150

151 ***OCA2* exon 10 skipping is the hallmark of primates and some mammalian species but not rodents**

152 All six single nucleotide substitutions tested above together with the 2 synonymous variants previously
153 described are evenly distributed along the 72 bp of exon 10 and have a significant effect on skipping
154 despite being distant more than 2bp away from the intron/exon junction (Fig 1a).

155 The alternative transcript resulting from *OCA2* exon 10 skipping is reported only in some mammalian
156 species. We searched the sequence databases for *OCA2* exon 10 skipped transcripts in a series of
157 mammals and could identify the alternative skipped transcript in primates and representatives of
158 different mammalian groups (carnivores, laurasiatherian, artiodactyla, and glires) including the rabbit
159 but not rodents such as mouse or rat (Table S2). Alignments of the exon 10 sequences from selected
160 representatives are provided in Fig S2. Alignment of the human and murine exon 10 sequences
161 highlights 12 nucleotide differences in the 72 bp exonic sequence (Fig 2a). Most of these differences
162 are synonymous. Only two codons are different, at position p.360, with Ser(TCC) in human and
163 Ala(GCA) in the mouse, and at position p.370, with Ile(ATT) in human and Val(GTT) in the mouse. Of
164 note, (A) lies at position c.1065 which corresponds to the common European human variant
165 c.1065G>A.

166 The splicing profile of exon 10 was further analyzed by RT-PCR on C57BL/6J mouse skin RNAs and
167 compared to the human splicing profile. Only one form of RT-PCR product including exon 10 was
168 obtained from mouse tissue, confirming the total retention of exon 10 in this rodent species (Fig 2b).
169 The absence of detectable exon 10 skipping was also verified by testing RNAs extracted from mouse
170 skin at different stages as well as pigmented optic cup samples (Fig S3a).

171 These findings prompted us to test the murine sequence in the minigene assay. As shown in Fig 2c, RT-
172 PCR performed on RNAs extracted from cells transfected with the murine sequence homologous to
173 the human construct (exon 10 surrounded by 335 bp of intron 9 and 313 bp of intron 10), only amplifies
174 a single band corresponding to the correctly spliced RNA with exon 10. The mouse minigene construct

175 therefore contains the minimal sequences for total exon retention and recapitulates the murine *in vivo*
176 profile.

177 We then used the human and the murine minigene constructs to test reciprocal exonic mutations as
178 summarized in Fig 2d (see Fig S3b-c for experimental data). Three codons of exon 10 were specifically
179 targeted. The first codon p.349 encodes Ile in both human and mouse with yet a different third
180 nucleotide, ATC/T (C in human, T in mouse). The human synonymous variant c.1047C>T/p.Ile349=
181 (Fig 1a) therefore corresponds to the murine codon (ATT). Yet, in the human context, it significantly
182 increased skipping as previously shown (12), whereas no skipping was detected with the reciprocal
183 change, ie. (ATC) in the murine context. This indicates that the effect of this nucleotide change is
184 dependent upon the surrounding sequences. Codon sequence at position p.360 was reciprocally
185 swapped. At this position the murine codon Ala(GCA) replacing Ser(TCC) in the human minigene
186 resulted in a marked increase in skipping whereas the reciprocal sequence did not induce skipping of
187 the murine exon 10. None of the nucleotide changes that were introduced in this codon resulted in
188 preventing skipping of the human minigene or induced skipping of the murine minigene. Swapping of
189 the codon sequence at p.370 with Val(GTT) in the human minigene or Ile(ATT) in the mouse minigene
190 had no effect, nor did the introduction of both human p.360 and p.370 sequences in the murine
191 minigene.

192

193 **Vulnerability of human *OCA2* to exon 10 skipping relies on the combination of exonic and intronic**
194 **sequences**

195 As none of the single nucleotide changes that we tested in the human exon sequence prevented its
196 skipping, we next questioned the role of the surrounding *OCA2* introns by assaying chimeric constructs
197 made of human and murine sequences (Fig 3). As illustrated in Fig 3a, when the human exon 10 was
198 flanked by mouse intron 9 and 10 sequences (~300 bp each), full retention was observed as with the
199 murine control minigene. The reciprocal construct (murine exon 10 flanked by human intronic

200 sequences) also resulted in full retention. The presence of only one of the segments (intron 9 or exon
201 10 or intron 10) of murine origin was sufficient to protect exon 10 from skipping. Moreover, in the
202 murine minigene, the intronic segments could be shortened down to 75 bp for intron 9 (including the
203 putative branch point at -37 and the polypyrimidine track) and 20 bp for intron 10 without affecting
204 exon 10 retention (Fig 3b). These results indicated that human exon 10 vulnerability is dependent upon
205 a combination of intronic and exonic sequences.

206 We next investigated in more detail the potential role of nucleotide at position c.1045-9 which is the
207 first divergent nucleotide between the human and murine sequences upwards the intron 9/exon 10
208 junction (Fig 2a). Interestingly at this position, (T) is associated to basal levels of skipping (human
209 reference sequence), whereas (G) is pathogenic as it causes full skipping as previously shown (12).
210 Moreover, the mouse sequence carries a (C) and is associated to full retention in the murine context.
211 Nucleotide c.1045-9G was introduced in the murine and chimeric constructs. As shown in Fig 3c, this
212 variant induced skipping only when embedded in an intron 9 sequence of human origin.

213 Sequence homology in this region was next analyzed in a series of mammalian species which showed
214 identity of the intron 9 sequence from c.1045-18 to the intron/exon junction among non-human
215 primates that express the skipped transcript (Table S2, Fig S2a). Among non-primate mammalian
216 species, only the rodent *Psammomys obesus* was found to have a 17 bp identical sequence in this
217 region (Fig S4a). *Psammomys obesus* (also known as Fat Sand Rat) is a diurnal gerbil with a pale coat
218 colour that has proved a valuable model for studying several eye diseases such as diabetic retinopathy
219 and Stargardts disease (26,27). The unexpected *OCA2* intron 9 sequence homology between this
220 rodent and primates prompted us to look at the splicing of its exon 10. RNAs were purified from skin
221 and pigmented optic cups of adults and used for RT-PCRs (Fig S4b). This showed that exon 10 is fully
222 retained in the *Oca2* mRNAs of *Psammomys* despite a much higher homology to human than mouse
223 or rat in the flanking intron 9 sequence.

224 The sequence corresponding to the shortened mouse minigene was next examined in selected
225 mammalian species that do (human and non-human primates, lion, blue whale, donkey, rabbit) or do
226 not allow exon 10 skipping (mouse, rat, *Psammomys*) (Table S2; Fig S2). This revealed a stretch of
227 ~40bp, -20 to -60 bp upstream of the intron 9 3'splice site of species vulnerable to exon 10 skipping
228 including the rabbit, which is absent in mouse, rat and *Psammomys*. In addition, alignment of the 20bp
229 of intron 10 included in the shortened mouse minigene showed that the sequence (+1 to +8)
230 recognized by the U1snRNA (28) is identical in all studied species whether exon 10 is vulnerable to
231 skipping or not. From +9 to +20, the sequence is identical in all rodents (in which exon 10 is
232 systematically included) but substantially variable among species vulnerable to exon 10 skipping.

233 In order to get more insight on the involvement of the intronic surrounding sequences, new human
234 minigene constructs were designed with shortened *OCA2* intronic segments (Fig 3d). Similar to the
235 murine construct (Fig 3b), the intron 9 sequence could be reduced to 73 bp (including the putative
236 branch point at -47bp) without significantly affecting the skipping ratio. Of note this sequence contains
237 the 40 bp stretch that is missing in the rodent sequence. By contrast, shortening of human intron 10
238 to 20 bp induced marked exon 10 skipping when combined to >300 bp of intron 9 and full skipping
239 when both intron 9 and intron 10 sequences were shortened to 73 and 28 bp respectively. Additional
240 constructs defined a minimal human minigene with preserved basal levels of skipping of less than 30%.
241 This construct carried 73 bp of intron 9 and 330 bp of intron 10 (Fig 3d). This highlighted a major
242 difference with the murine minimal sequence that requires only 20 bp of intron 10 to fully retain the
243 exon in the minigene assay.

244 This study was extended by using antisense RNA oligonucleotide (ASO) mapping to investigate more
245 precisely the role of different regions of exon 10 in retention/skipping control. Four ASO sequences of
246 22 nucleotides each were synthesized to walk along the entire 72 bp exon 10 (Fig 4a). They were
247 designed to hybridize the human sequence and were used to co-transfect HeLa cells with the human
248 minigene at concentrations of 50 or 5 nM. At 50 nM, each of ASO1, ASO2 and ASO4 induced complete

249 skipping. ASO3 allowed a few percent of RT-PCR products to retain exon 10 (Fig 4b). The involvement
250 of the four regions covered by the ASOs was confirmed using them at lower concentration of 5nM
251 (Fig 4c). ASO3 allowed retention of exon 10 in less than 30% of the transcripts whereas the 3 other
252 ASOs still induced major skipping (~90%). We then used the same assay with the chimeric construct
253 containing the human exon 10 flanked by the murine intronic sequences. Using this chimeric minigene,
254 a concentration 50 nM of ASO (Fig 4d) was required to achieve a similar skipping profile to that
255 observed with the human minigene sequence using 5nM ASO. Interestingly in the presence of murine
256 intronic sequences and a concentration of 5nM of ASO, none of the ASOs were able to substantially
257 induce exon 10 skipping which was detected in only a small and variable fraction of the RT-PCR
258 products after treatment with ASO1, 2 and 4 (Fig 4e). These results confirmed that exon 10 retention
259 skipping balance relies on both the exon sequence and its intronic environment suggesting that co-
260 evolution of these segments allowed basal levels of skipping in several groups of mammalian species
261 but not rodents. This hypothesis is supported by the comparison of the nucleotide sequences across
262 the whole region in species that do or do not express the skipped transcript (Fig S2).

263

264 **The benign variant (c.1065G>A/p.Ala355=) rs1800404-T enhances exon 10 skipping in the minigene**
265 **assay**

266 We next tested the most frequent exonic variant of *OCA2* that locates in exon 10. The synonymous
267 c.1065G>A/p.Ala355= corresponds to the (T) allele of SNV rs1800404:C/T. It is highly predominant in
268 European populations with an average allelic frequency of around 0.8 whereas it is less represented in
269 Asian (f= ~0.4) and African, African American populations (f= ~0.2) as extracted from the GnomAd
270 database (29). Interestingly, statistical association studies using RNA-Seq data obtained on a collection
271 of primary melanocytes highlighted a strong correlation between c.1065G>A (referred to as
272 rs1800404-T in population genetic studies and here after) and exon 10 skipping (18). As all of our
273 patients with European ancestry carry c.1065G>A in *cis* to their *OCA2* exon 10 VUS, our minigene assay

274 had so far been performed on a c.1065G>A background, which, as illustrated in Fig 1c, resulted in ~17%
275 basal skipping levels in the WT control. In order to monitor the potential impact of c.1065G>A on exon
276 10 skipping, a control minigene carrying c.1065G was compared with our standard control c.1065G>A
277 control minigene in the splicing assay. In addition, c.1080C>T/p.Ser360= previously shown pathogenic
278 in *cis* to c.1065G>A (12) was comparatively tested in *cis* to either c.1065G or c.1065G>A. As shown in
279 Fig 5, c.1065G allowed a significantly better retention of exon 10 (~4.5-fold) with basal skipping levels
280 of only ~3%. The enhancement of skipping driven by c.1080C>T/p.Ser360= was significant in *cis* to
281 c.1065G with an increase from 3% to 25% of skipped RT-PCR products. This level was however much
282 weaker than in *cis* to c.1065G>A (65%). The additive effect of c.1065G>A was also observed in *cis* to
283 the other previously studied rare synonymous variant c.1047C>T/p.Ile349= (12) as shown in Fig S5.

284

285 **rs1800404-T is associated with lighter pigmentation in European population**

286 The c.1065G>A variant significantly enhances skipping of exon 10 including when located in *cis* to a
287 rare variant such as c.1080C>T, which is likely to contribute to reaching the threshold for pathogenicity.
288 In the absence of a rare variant in *cis*, it induces a 4.5-fold skipping increase compared to c.1065G in
289 the minigene assay. This raises the question of its functional impact on pigmentation in individuals
290 without albinism, particularly in the European population for which this SNV has never been associated
291 with pigmentation scores. Of note, rs1800404:C/T is located at ~5kb of rs1800407:C/T of which the
292 derived allele rs1800407-T (corresponding to c.1256G>A/p.R419Q) is reported as likely hypomorphic
293 with suggested causative effect in lighter pigmentation (30). In European populations, rs1800404 and
294 rs1800407 are in significant linkage disequilibrium (LD, r^2: 0.27; D': +0.95) as shown by analysing the
295 LDlink database (31). This pattern of LD generates 3 common haplotypes, rs1800404-T/rs1800407-C
296 (f= 0.783), rs1800404-C/rs1800407-C (f= 0.141) and rs1800404-C/rs1800407-T (f= 0.073), whereas
297 haplotype rs1800404-T/rs1800404-T is much less represented (f= 0.003). We hypothesized that the

strong LD between these SNVs may hide the association of rs1800404:C/T to pigmentation scores in European populations with a majority of fair skin.

For this reason, we investigated the association of haplotypes derived from rs1800404 and rs1800407 to pigmentation phenotypes in a UK Biobank cohort with European reported ancestry (n=402871). Fig 6a-b shows the distribution of the different self-reported skin colours (6 categories) and hair colours (6 categories) in this population as well as of the genotypes at rs1800404 (f(T)= 0.792) and rs1800407 (f(T)= 0.086), illustrating the under-representation of haplotype rs1800407-T/rs1800404-T with only 2 double homozygotes (f=5 x 10^{-6}) carrying the derived allele of each SNV (Fig 6c). In the selected cohort, the pattern of LD between rs1800404 and rs1800407 was consistent (r^2 = 0.33; D' = + 0.96) with that reported in the European population of the LDlink database.

Taking the most represented phenotype as a reference, ie. fair for the skin, and light brown for the hair colour, we then investigated the association of rs1800407-T and rs1800404-T alleles with skin and hair pigmentation using a multinomial regression model.

Under the assumption of proportional odds, single SNV analysis showed that the rs1800407-T allele was associated with lighter skin pigmentation (OR = 0.916 [0.901 - 0.932], p = 4.83 x 10^{-23}) as was the rs1800404-T allele (OR = 0.924 [0.914 - 0.935], p = 1.12 x 10^{-38}), Fig S6a and Table S3a. In a joint model assuming additive effects of the two T alleles, these associations were significantly reinforced, OR = 0.798 [0.781 - 0.815] (p = 1.04 x 10^{-98}) and OR = 0.846 [0.834 - 0.858] (p = 1.99 x 10^{-115}), respectively. Detailed results of the haplotype analysis of these two SNVs are provided in Table S3a and are summarized in Fig 6d. Compared to the rs1800404-C/rs1800407-C haplotype, all three other haplotypes were associated with lighter skin pigmentation: rs1800404-T/rs1800407-C: OR = 0.853 [0.841 – 0.866] (p = 2.05 x 10^{-101}); rs1800404-C/rs1800407-T: OR = 0.813 [0.796 – 0.831] (p = 1.10 x 10^{-78}] and rs1800404-T/rs1800407-T: OR = 0.472 [0.426 – 0.523] (p = 2.56 x 10^{-46}). This illustrates the strong correlation of rs1800404-T to fair pigmentation, an association that can go undetected if LD between rs1800404 and rs1800407 is not taken into account. In turn, the impact of rs1800407-T on

323 skin pigmentation tended to be significantly ($p = 7.18 \times 10^{-13}$) higher in *cis* to rs1800404-T allele (OR =
324 0.553) rather than to rs1800404-C allele (OR = 0.813).

325 Similar analyses were next conducted on hair pigmentation (Fig 6e, Table S3b; Fig S6c-d). The cohort
326 was reduced to the four major hair colour categories (n=379,183 subjects) excluding subjects with red
327 or undefined hair colour. Single SNV analysis unexpectedly found that rs1800407-T was associated to
328 dark hair colour rather than light (OR = 1.432[1.410 - 1.454] $p < 10^{-306}$) whereas rs1800404-T was, as
329 expected under the assumption of a causal role, significantly correlated to fair hair (OR = 0.819 [0.810
330 - 0.828], $p = 1.21 \times 10^{-306}$). In a joint additive model, these opposite directions were confirmed with
331 rs1800407-T associated with increased hair pigmentation (OR = 1.334 [1.310 – 1.359] ($p = 4.35 \times 10^{-}$
332 ^{205}) and rs1800404-T allele associated with decreased hair pigmentation (OR = 0.919 [0.907 - 0.931], p
333 $= 8.00 \times 10^{-39}$). Haplotype analysis further demonstrated that the rs1800404-T/rs1800407-C and
334 rs1800404-T/rs1800407-T haplotypes were associated with lighter hair pigmentation (OR = 0.932[0.92
335 - 0.944], $p = 7.24 \times 10^{-27}$) and OR = 0.692[0.633 - 0.756], $p = 3.44 \times 10^{-16}$, respectively, compared to the
336 rs1800404-C/rs1800407-C haplotype. By contrast, the rs1800404-C/rs1800407-T was markedly
337 associated with darker hair pigmentation (OR = 1.376[1.35 - 1.402], $p = 7.99 \times 10^{-236}$). These
338 observations demonstrate that considering the haplotype rs1800404/rs1800407 is key in statistical
339 association analyses with pigmentation-related phenotypes.

340 Furthermore, these results support the hypothesis that the increase of exon 10 skipping triggered by
341 rs1800404-T negatively impacts pigmentation not only in the context of albinism but also in the whole
342 population.

343

344 **The alternative *OCA2* transcript with skipped exon 10 may be expressed as a protein isoform with**
345 **dominant-negative effect**

346 We next asked how a moderate increase in skipping, such as that mediated by rs1800404-T, could
347 significantly impact melanin production in the pigmented cells of healthy individuals. As exon 10 of

348 *OCA2* is in frame with the rest of the coding sequence, skipping it theoretically results in the translation
349 of an isoform lacking a stretch of 24 amino acids (from p.349 to p.372). If so, it raises questions about
350 its potential function. In an attempt to predict the consequences of increased levels of the putative
351 protein, which we will refer to as ΔE10-OCA2, we modelled and compared its topology to the topology
352 of the native OCA2 protein initially published and revised by some of us (15,32). This indicated a
353 substantial change in the arrangement of the first transmembrane helix domain that no longer packs
354 with the main transmembrane helical bundle (Fig 7a). This prediction and the associated lower
355 confidence scores (Fig S7a) indicate that ΔE10-OCA2 can fold with a likely lower structural stability than
356 that of OCA2. Furthermore, multimer modelling predicts that the OCA2 protein can assemble as a
357 homodimer (32). This raises the question of a potential dominant-negative effect of increased levels
358 of ΔE10-OCA2 resulting from increased skipping of exon 10. As shown in Fig 7b, multimer models of
359 the ΔE10-OCA2 homodimer and the heterodimer with full-length OCA2 protein are expected with
360 overall high confidence scores despite a reduced confidence at some predicted interface regions
361 (Fig S7b). All in all, these models predict that the formation of heterodimers between full-length OCA2
362 and ΔE10-OCA2 isoforms is biologically feasible. Further experimental investigations will be required
363 to functionally test the hypothesis of dominant-negative effect of ΔE10-OCA2 through
364 heterodimerization.

365

366 # Discussion

367 In this study, we provide evidence that the human *OCA2* gene exhibits high sensitivity to sequence
368 modifications within and surrounding exon 10, resulting in significant alterations of skipping levels that
369 impact pigmentation in both pathological and physiological contexts. Previously, we classified two
370 synonymous variants of exon 10 as pathogenic due to exon 10 skipping enhancement (12). Here, we
371 demonstrate that five out of six rare missense variants in exon 10 also increase exon skipping to varying
372 degrees, which can contribute to the attainment of the pathogenic threshold. Functional assays

373 combining murine and human sequences illustrate the involvement of the whole exon in exon
374 retention/skipping control. A comparable sensitivity to sequence-dependent modulation of skipping
375 levels has been reported and extensively studied in the case of *CFTR* exon 13 (historical name: exon
376 12) by Pagani's group (33–35). Like *OCA2* exon 10 (72 bp), exon 13 of *CFTR* (87 bp) is in-frame with the
377 upstream and downstream coding exons. Its skipping levels are strikingly dependent on multiple
378 nucleotides all along the exon, with many single nucleotide variants (missense or synonymous) being
379 pathogenic due to excessive skipping. Interestingly, the skipping level of *CFTR* exon 13 is much lower
380 in mouse compared to human and non-human primates, which allowed the authors to identify critical
381 exonic sequences involved in the splicing/skipping balance (34). This raised hypotheses of dual
382 selective pressures towards loss of exon retention and amino-acid sequence optimization in the course
383 of evolution (35). Although *OCA2* exon 10 has a similar profile as *CFTR* exon 13, ie basal levels of
384 skipping in primates and no detectable skipped transcripts in rodents, our observations are not in
385 favour of selection pressures towards codon or amino-acid optimization in species that do express
386 basal levels of skipped transcripts such as reported for *CFTR* (34) and other genes (36). Moreover, they
387 highlight the involvement of the intronic surrounding sequences that have coevolved in mammalian
388 species with basal levels of skipping. This leaves the possibility that the expression of the skipped
389 transcript has been positively selected through evolution for a functional role in pigmentation control
390 including within the human population like that proposed for other genetic variants of *OCA2* (37).

391 In line with this hypothesis, we show that the most frequent exonic human variant c.1065G>A
392 [rs1800404:C/T], which is predominant in European population with fair skin, significantly increases
393 exon 10 skipping. This result functionally supports the strong association between rs1800404-T and
394 exon 10 skipped transcript previously reported after RNA-seq analysis of biological samples (18). In this
395 same GWAS study that identified several loci associated with skin pigmentation in African populations,
396 rs1800404-T displayed the highest probability of being causal. Strong association was also reported in
397 several other population with dark skin (19–22) but not in European populations with lighter skin. We
398 hypothesized that this may be due to strong LD between rs1800404 and the previously identified

399 nearby SNV rs1800407. Indeed, our haplotype association analysis of over 400,000 individuals with
400 European ancestry shows that rs1800404-T, which causes exon 10 skipping enhancement, strongly
401 correlates with lighter skin and hair in Europeans whereas the direction of the effect of rs1800407 is
402 not concordant with causality of hair colour (Fig 6e and Fig S6c) as previously reported on a smaller
403 cohort (38). The association as well as the predominant frequency of rs1800404-T in the European
404 population (f= ~0.8) raise intriguing questions about its evolutionary selection such as its potential
405 additional role in the adaptation of human skin colour in different regions of the globe (39).

406 Our results together with previously reported findings (18) suggest that *OCA2* exon 10 skipping levels
407 induce proportional decrease in pigmentation in both pathological and physiological contexts. This
408 prompts us to consider the underlying mechanisms that could enable such a fine-tuned phenotypic
409 response. Although we have no evidence for the expression of the corresponding protein ΔE10-OCA2,
410 its potential dominant-negative effect would explain the dose-sensitive response suggested by
411 statistical associations in healthy populations. Given that the OCA2 protein may well function as a
412 dimeric transporter and shares homology with members of the NACT family (32,40) which includes
413 transporters with cooperative or interdependent subunit mechanisms, the presence of the non-
414 functional ΔE10-OCA2 monomer within a heterodimer could potentially compromise overall chloride
415 conductance, resulting in a dose-dependent down-regulation of melanogenesis. Alternatively,
416 heterodimer misfolding could induce endoplasmic reticulum-associated degradation (ERAD) via the
417 ubiquitin–proteasome system, similar to the quality control that is triggered by the well-studied ΔF508
418 form of the CF chloride conductance protein (41). Thus, whether engaged in a dimer or as a structurally
419 modified monomer, ΔE10-OCA2 could exert a dominant effect by interfering with *OCA2*-dependent
420 Tyrosinase processing and/or activity that is key to melanogenesis activation (8).

421 What are the implications of the present study for molecular diagnosis of albinism? Oculocutaneous
422 albinism is considered as a classical recessive rare disease caused by Mendelian rare variants with
423 strong functional effects as opposed to common genetic diseases caused by multiple common variants

424 with weak effects. We have recently provided several examples of complex determinism of albinism
425 that challenge this paradigm (42–44). Here, we show that rare variants of *OCA2* exon 10 such as
426 c.1047C>T and c.1080C>T that do enhance skipping without inducing full exclusion, have a much
427 weaker effect in *cis* to c.1065G compared to c.1065G>A, questioning the role of the haplotype in the
428 refinement of pathogenicity threshold. We therefore propose that any newly identified VUS in the
429 exon 10 region should be tested in a minigene assay taking into account the haplotypic combination
430 with c.1065 carried by the patient. All variants causing skipping enhancement that were identified in
431 our previous report (c.1047C>T/p.Ile349= ; c.1080C>T ; p.Ser360= ; c.1095_1103del/p.Ala366_368del)
432 or in this study (c.1048G>A/p.Val350Met ; c.1081C>G/p.Leu361Val) were in *cis* to c.1065G>A in the
433 corresponding patients of our local cohort. Yet, their impact in *cis* to c.1065G is unproven. Interestingly,
434 four of the six missense variants that were studied here have a significantly more elevated frequency
435 in some non-European populations than worldwide (Table S4). For instance c.1048G>A/p.Val350Met
436 is significantly more frequent in South Asia ($f=1.3.10^{-3}$) than in the general population ($f=9.79.10^{-5}$).
437 Two presumably non-albino homozygotes carrying this variant are scored in the South-Asian
438 population. Their exome sequence confirms that none of the corresponding alleles have the
439 c.1065G>A (f= 0.39 in South Asia) in *cis* to c.1048G>A/p.Val350Met (Fig S8). The same observation can
440 be made with the three other missense variants shown here to significantly increase skipping in *cis* to
441 c.1065G>A (c.1064C>T/p.Ala355Val ; c.1109T>C/p.Ile370Val ; c.1114G>A/p.Asp372Asn, Fig 1c).
442 Sequencing data indicate that they systematically segregate with c.1065G in presumably healthy
443 homozygotes (Fig S8) leaving the possibility that the pathogenic threshold is reached only when they
444 lie in *cis* to c.1065G>A. In the context of future diagnosis, these observations suggest that any such
445 variant may impact pigmentation within the physiological range if located in *cis* to c.1065G and be
446 pathogenic in *cis* to 1065A. Not only c.1065 (rs1800404) but the entire haplotype that regulates *OCA2*
447 expression should be analyzed to enable the prediction of *cis*-interfering or *cis*-additive effects of
448 common variants. This includes the enhancer of *OCA2* within *HERC2* and its causal polymorphism
449 rs12913832 (5). As a result, the identification of haplotypes with combination of SNVs that result in

450 high skipping ratio may resolve the genetic determinism of some of the patients with so far only one
451 single identified mendelian variant in *OCA2*. In the light of the new findings presented here, several
452 already documented variants with inconclusive or conflicting interpretation would be worth
453 reconsidering. For instance, the synonymous variant c.1113C>T/p.Gly371= (rs1800405-A) is reported
454 as likely causal in a family of Vietnamese patients (45) whereas it is considered as benign due to its
455 frequency in the range of the percent worldwide reaching ~10% in Admixed American populations with
456 as many as 1% of homozygotes (GnomAd database). Of note, linkage disequilibrium with c.1065
457 (rs1800404) makes the haplotype rs1800405-A/rs1800404-T absent from the 5008 haplotypes
458 included in the LDlink database. Testing of the 4 possible haplotypes in the minigene assay together
459 with haplotype identification in the corresponding patients should allow to resolve their genetic
460 diagnosis. This strategy applies to other variants listed in Table S4.

461 As summarized in Figure 8, this study strongly suggests that c.1065G>A (rs1800404-T) contributes to
462 paleness/hypopigmentation on the whole spectrum from healthy population with dark skin to patients
463 with albinism. Including rs1800404 in genetic testing not only applies to OCA2 patients with missing
464 heritability but also to patients with other types of albinism taking into account epistasis of *OCA2* on
465 downstream genes. Whether this polymorphism is also associated with specific retinal phenotypes in
466 albino and non-albino individuals merits future investigations as these findings could have direct
467 implications for refining genetic diagnosis of albinism.

468

469 # Materials and Methods

470 **Ethics statement**

471 This study was approved by our local ethics committee: Comité de Protection des Personnes
472 Bordeaux Outre-Mer III. Written informed consent was received from the patients and the
473 participating families. The UK Biobank has received approval from the National Information
474 Governance Board for Health and Social Care and the National Health Service North West Centre for
475 Research Ethics Committee (Ref: 11/NW/0382). All investigations were conducted in accordance with
476 the tenets of the Declaration of Helsinki.

477 **Recombinant plasmid constructs and site-directed mutagenesis**

478 The minigene containing the human wild-type (WT) *OCA2* exon 10 in pSPL3B vector has been described
479 (12). For constructing the minigene with the murine *Oca2* sequences, genomic DNA was extracted
480 from C57BL/6J mouse tail-tip using "Phire Tissue Direct PCR Master Mix" (Thermo Fisher Scientific)
481 following manufacturer's recommendations. Exon 10 flanked by 335 bp of intron 9 and 313 bp of
482 intron 10 were amplified with "Q5 High Fidelity PCR Kit" (New England Biolabs) according to
483 manufacturer's instructions and cloned in pSPL3B vector by homologous recombination as previously
484 described (12). PCR product was treated with the "In-Fusion HD Cloning kit with Cloning Enhancer
485 Treatment" (Takara Bio) and recombined with Not I linearized pSPL3B vector. Both human and mouse
486 minigenes served as a template to generate constructs with shortened introns. For the human
487 minigene constructs, sequence of intron 9 was reduced to 73bp that include the putative branch point.
488 Four versions with different amounts of intron 10 sequences were cloned (330 bp, 212 bp, 85 bp or 28
489 bp) as well as combinations of shortened intron 9 (73 bp) and intron 10 (429 bp) or intron 9 (326 bp)
490 and intron 10 (28 bp). The mouse minimal minigene was shortened to 75bp for intron 9 and 20bp for
491 intron 10. All chimeric human/mouse constructs containing exon 10 flanked by around ~400 bp of
492 intron 9 and 10 were synthesized by Eurofins–Genestrand. Lyophilized DNA was resuspended in

493 nuclease-free water, cloned in pMiniT vector using the “PCR cloning kit” (New England Biolabs). The
494 inserts were amplified as described above. PCR products were extracted from 1% agarose gel and
495 purified using the “NucleoSpin Gel and PCR Clean-up kit” (Macherey-Nagel) and cloned in pSPL3B
496 vector as above. All constructs were confirmed by Sanger sequencing (Eurofins). All primers that were
497 used in this study are listed in Supplementary Table S1. For site-directed mutagenesis, single nucleotide
498 variants of interest were introduced into the reference sequence of the minigene with either standard
499 “QuickChange Site-Directed Mutagenesis Kit” or with optimized “QuickChange Lightning Site-Directed
500 Mutagenesis Kit” (Agilent) following the manufacturer's recommendations. Mutagenesis primers were
501 designed as recommended and are listed in Table S1. Targeted mutagenesis was confirmed by Sanger
502 sequencing (Eurofins).

503 **RT-PCR on RNA from cultured cells and tissues**

504 All animal tissue samples used in this study were acquired from left-over post-mortem tissues from
505 other studies involving wild-type control animals (mouse and Fat Sand Rat) approved by Bordeaux
506 University Ethics Committee and Strasbourg CREMEAS. Total RNAs from one-month C57BL6/J mouse
507 skin, as well as skin and optic cups of Fat Sand Rat *Psammomys obesus*, were purified with “RNeasy
508 Mini kit” (Qiagen). Total RNAs from human primary melanocyte culture were obtained as described
509 (12). First-strand cDNA synthesis was carried out with 1µg of total RNAs by oligo(T)-primed reverse-
510 transcription using M-MLV reverse transcriptase (Promega) following the manufacturer’s instructions.
511 The resulting cDNAs were amplified by PCR with “Q5 High Fidelity PCR Kit” (New England Biolabs)
512 containing human (forward: 5’ CTCCGCGGAAGTGTAGAAAC 3’ and reverse: 5’
513 GTGAAGAGGAGCATGGTGGT 3’) or mouse (forward: 5’ CTCTCCAACAGACTCGCCTC 3’ and reverse: 5’
514 AGGAGCATCGTAGTGACATTGT 3’) primers located in exon 9 and 13, or *Psammomys obesus* primers
515 located in exon 8 and 12 (forward: 5’ CCTCCAACAGACCCACCTTG 3’; reverse: 5’
516 GTGGTGACATTGTCTAAGAAGGC 3’). PCR products were separated by electrophoresis through 1.5%
517 agarose gel, verified by Sanger sequencing (Eurofins) and aligned with reference *OCA2* transcripts

518 sequences (human: ENST00000354638.8, mouse: ENSMUST00000032633.12, www.ensembl.org),
519 (*Psammomys obesus*: XM_055607524.1, www.ncbi.nlm.nih.gov).

520 **Antisense RNA oligonucleotide (ASO) synthesis**

521 ASOs were designed to map on human *OCA2* exon 10 with the following sequences: ASO1 : 5'-
522 CUGCCAGAGUUCUGUGCACGAU-3' ; ASO2 : 5'-GCAAGGGAACCCAGCAUGGCUG-3' ; ASO3 : 5'-
523 CCAGUGCUGCCAGUGCUGCAAG-3'; ASO4 : 5'-AUCGCCAAUCACAGCCAGUGCU-3'. A non-targeting RNA
524 oligonucleotide was used as a negative control (5'- AGUUCACACUGGAAUCAAAGGU-3'). ASOs were
525 synthesized on a micromolar scale by the β-phosphoramidite method with backbone modifications
526 (phosphothioate and 2'-methoxyethyl) using an H8 automated synthesizer (K&A Labs, Germany).
527 Deprotection of the ASOs was performed according to the supplier's protocol. The concentrated crude
528 ASOs were then resuspended in water. The sample concentration was determined from the
529 absorbance at 260 nm and molar extinction coefficient of the ASO. This value was calculated using the
530 Integrated DNA Technology online oligo analyzer tool, which uses the standard nearest neighbor
531 method.

532 **Cell culture and transfection**

533 HeLa cells were cultured in DMEM+GlutaMAX™ (Gibco) supplemented with 1% penicillin and
534 streptomycin and 10% fetal bovine serum (Gibco). The day before transfection HeLa cells were seeded
535 in 6-well plates with 150 000 cells per well. For minigene assays, transfections of HeLa cells were
536 performed using FuGENE® HD Transfection Reagent (Promega) with 3µg of plasmid DNA in a 1:1 ratio
537 according to the manufacturer's recommendations. Cells were collected forty-eight hours after
538 transfection by scraping. For ASO mapping experiment, HeLa cells were transfected with
539 Lipofectamine3000 (Thermo Fisher Scientific) with 1µg of plasmid DNA and 5 or 50nM of ASO.
540 Optimization of the standard protocol provided by the supplier consisted in increasing the mix
541 incubation time from 5 to 20 minutes. To improve cell transfection efficiency, the transfection mix was
542 applied to the cells in Opti-MEM 1X (Gibco). The cells were left in Opti-MEM at 37°C for 4 hours after

543 which the Opti-MEM was replaced by complete culture medium. Cells were collected the day after
544 transfection.

545 **Minigene products analysis**

546 Total RNAs were purified using the "RNeasy Mini kit" (Qiagen). First-strand cDNA synthesis was carried
547 out as described above. The resulting cDNA was amplified by PCR using SD6 and SA2 vector specific
548 primers (SD6: 5' TCTGAGTCACCTGGACAACC 3'; SA2: 5' ATCTCAGTGGTATTTGTGAGC 3') as described
549 before (12). PCR products were separated by electrophoresis through 2% agarose gel. In order to
550 quantify *OCA2* exon 10 skipping, PCR products were first purified according to "NucleoSpin Gel and
551 PCR Clean-up" (Macherey-Nagel) instructions and separated on a D1000 DNA Screentape system with
552 TapeStation 4150 instrument (Agilent) following manufacturer's recommendations. The percentage
553 integrated area (PIA) of the peak corresponding to the RT-PCR product without exon 10 is directly used
554 to estimate the % of exon skipping.

555 **UK Biobank cohort**

556 The UK Biobank is a large well-studied prospective cohort with more than 500,000 participants
557 recruited from across the UK. Participants completed a baseline questionnaire including self-reporting
558 of their skin and hair colour (46). Participants provided extensive health, lifestyle, and genetic data,
559 including self-reported traits such as skin and hair color, captured during baseline assessment visits.
560 Genotyping was performed using custom UK Biobank Axiom arrays, and genotype imputation was
561 conducted centrally using the Haplotype Reference Consortium and UK10K reference panels. Variant-
562 level data were extracted from the imputed dataset using PLINK v2.0 by specifying genomic
563 coordinates or rsIDs of interest. Relevant pigmentation-related traits were extracted using the DNA
564 Nexus cohort browser and Table Exporter v2.0.82.

565 **OCA2 protein model building and dimer prediction**

566 OCA2 monomer structures were generated using a local installation of ColabFold (47). Dimeric models
567 were constructed using the AlphaPulldown pipeline (48), which incorporates the AlphaFold-Multimer
568 framework (49,50). For both monomeric and dimeric predictions, multiple sequence alignments were
569 obtained using MMseqs2 (51). For each of the three dimer configurations, five structural models were
570 generated, and the top-ranked model based on the internal scoring metric was selected for
571 downstream analysis.

572 **Statistical analysis**

573 Functional assays: all the results are expressed as mean value ± standard error of the mean (SEM) of
574 at least 3 independent experiments, unless stated otherwise. Statistical analysis was performed with
575 GraphPad Prism version 8.0.2 for Windows using one-way ANOVA followed by the Bonferroni post hoc
576 multiple comparison test.

577 Population genetics analysis: *OCA2* haplotypes derived from rs1800404:C/T and rs1800407:C/T were
578 inferred using the *haplo.stats* R package. Associations between OCA2 SNVs/haplotypes and skin and
579 hair pigmentation were analyzed using polytomous regression models (including both multinomial and
580 proportional odds models), adjusting for age and sex. The most common phenotype was used as the
581 reference category: fair skin and light brown hair, respectively. Regression analyses were conducted
582 using the *vglm* function from the *VGAM* R package. Haplotypes with estimated frequencies below
583 0.001 in the UK Biobank data were excluded from statistical association analyses.

584

585 # Acknowledgments

586 The authors would like to thank Genespoir, the French albinism association, for its invaluable support
587 and the significant help it provides to patients.

588 # Funding

589 We acknowledge the following sources of funding: Genespoir, the French albinism association; the
590 French National Research Agency (Agence Nationale de la Recherche; ANR-21-CE17-0041-01); the
591 Wellcome Trust (224643/Z/21/Z, Clinical Research Career Development Fellowship to P.I.S.); the UK
592 National Institute for Health Research (NIHR) Clinical Lecturer Programme (CL-2017-06-001 to P.I.S.);
593 the NIHR Manchester Biomedical Research Centre (NIHR 203308 to P.I.S.).

594 # Data availability statement

595 Novel variants have been uploaded to ClinVar public database.

596 Data for building Figs 1, 3, 4, 5 and related statistical analysis can be accessed at
597 https://doi.org/10.5281/zenodo.15829093

598 AlphaPulldown output files for the OCA2 and ΔE10-OCA2 can be accessed at
599 https://doi.org/10.5281/zenodo.15828886

600 UK Biobank data are available under restricted access through a procedure described at
601 http://www.ukbiobank.ac.uk/using-the-resource/. This research was conducted using the UK Biobank
602 Resource under project 53114.

603 All other data supporting the findings of this study are available within the article (including its
604 Supplementary Information files).

605

606 **AUTHOR CONTRIBUTIONS**

607 Conceptualization: SJ, BA, PIS; Formal Analysis: DAT, SM, DJR, SJ, EM; Investigation: EM, SC, VM, DAT,
608 OLG, OJC, SM, DJG; Methodology: EM, SC, DAT, DJG, SM, CEA; Project Administration: SJ, BA;
609 Resources: DH, PIS, DJR, VM; Supervision: SJ, Validation: SJ, Visualization: EM, SJ, SC, DAT, SM, Writing:
610 SJ, EM

611

612 # References

613

614 1. Bakker R, Wagstaff EL, Kruijt CC, Emri E, van Karnebeek CDM, Hoffmann MB, et al. The retinal
615 pigmentation pathway in human albinism: Not so black and white. Prog Retin Eye Res. nov
616 2022;91:101091.

617 2. Kruijt CC, de Wit GC, Bergen AA, Florijn RJ, Schalij-Delfos NE, van Genderen MM. The Phenotypic
618 Spectrum of Albinism. Ophthalmology. 1 déc 2018;125(12):1953-60.

619 3. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M, et al. The HIrisPlex-S
620 system for eye, hair and skin colour prediction from DNA: Introduction and forensic
621 developmental validation. Forensic Sci Int Genet. juill 2018;35:123-35.

622 4. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, et al. Global skin
623 colour prediction from DNA. Hum Genet. 1 juill 2017;136(7):847-63.

624 5. Visser M, Kayser M, Grosveld F, Palstra RJ. Genetic variation in regulatory DNA elements: the
625 case of OCA2 transcriptional regulation. Pigment Cell Melanoma Res. 2014;27(2):169-77.

626 6. Bellono NW, Escobar IE, Lefkovith AJ, Marks MS, Oancea E. An intracellular anion channel critical
627 for pigmentation. Miller C, éditeur. eLife. 16 déc 2014;3:e04543.

628 7. George A, Sharma R, Pfister T, Abu-Asab M, Hotaling N, Bose D, et al. In vitro disease modeling of
629 oculocutaneous albinism type 1 and 2 using human induced pluripotent stem cell-derived retinal
630 pigment epithelium. Stem Cell Rep. 11 janv 2022;17(1):173-86.

631 8. Ni-Komatsu L, Orlow SJ. Heterologous expression of tyrosinase recapitulates the misprocessing
632 and mistrafficking in oculocutaneous albinism type 2: Effects of altering intracellular pH and
633 pink-eyed dilution gene expression. Exp Eye Res. 1 mars 2006;82(3):519-28.

634 9. Lavado A, Jeffery G, Tovar V, de la Villa P, Montoliu L. Ectopic expression of tyrosine hydroxylase
635 in the pigmented epithelium rescues the retinal abnormalities and visual function common in
636 albinos in the absence of melanin. J Neurochem. févr 2006;96(4):1201-11.

637 10. Roffler-Tarlov S, Liu JH, Naumova EN, Bernal-Ayala MM, Mason CA. L-Dopa and the albino riddle:
638 content of L-Dopa in the developing retina of pigmented and albino mice. PloS One.
639 2013;8(3):e57184.

640 11. Lasseaux E, Plaisant C, Michaud V, Pennamen P, Trimouille A, Gaston L, et al. Molecular
641 characterization of a series of 990 index patients with albinism. Pigment Cell Melanoma Res. juill
642 2018;31(4):466-74.

643 12. Michaud V, Sequeira A, Mercier E, Lasseaux E, Plaisant C, Hadj-Rabia S, et al. Unsuspected
644 consequences of synonymous and missense variants in OCA2 can be detected in blood cell RNA
645 samples of patients with albinism. Pigment Cell Melanoma Res. sept 2024;37(5):534-45.

646 13. Holm LL, Doktor TK, Flugt KK, Petersen USS, Petersen R, Andresen BS. All exons are not created
647 equal—exon vulnerability determines the effect of exonic mutations on splicing. Nucleic Acids
648 Res. 8 mai 2024;52(8):4588-603.

649 14. Baralle M, Baralle FE. The splicing code. Biosystems. févr 2018;164:39-48.

650 15. Diallo M, Courdier C, Mercier E, Sequeira A, Defay-Stinat A, Plaisant C, et al. Functional
651 Characterization of Splice Variants in the Diagnosis of Albinism. Int J Mol Sci. 8 août
652 2024;25(16):8657.

653 16. Diallo M, Sylla O, Sidibé MK, Plaisant C, Mercier E, Sequeira A, et al. Genotypic spectrum of
654 albinism in Mali. Pigment Cell Melanoma Res. nov 2024;37(6):752-61.

655 17. Sitaram A, Piccirillo R, Palmisano I, Harper DC, Dell'Angelica EC, Schiaffino MV, et al. Localization
656 to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2
657 function. Mol Biol Cell. mars 2009;20(5):1464-77.

658 18. Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, et al. Loci associated with
659 skin pigmentation identified in African populations. Science. 17 nov 2017;358(6365):eaan8433.

660 19. Adhikari K, Mendoza-Revilla J, Sohail A, Fuentes-Guajardo M, Lampert J, Chacón-Duque JC, et al.
661 A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in
662 Eurasia. Nat Commun. 21 janv 2019;10(1):358.

663 20. Anno S, Abe T, Yamamoto T. Interactions between SNP Alleles at Multiple Loci Contribute to Skin
664 Color Differences between Caucasoid and Mongoloid Subjects. Int J Biol Sci. 2008;81-6.

665 21. Batai K, Cui Z, Arora A, Shah-Williams E, Hernandez W, Ruden M, et al. Genetic loci associated
666 with skin pigmentation in African Americans and their effects on vitamin D deficiency. PLoS
667 Genet. févr 2021;17(2):e1009319.

668 22. Rivera-Paredez B, Hidalgo-Bravo A, López-Montoya P, Becerra-Cervera A, Patiño N, Denova-
669 Gutiérrez E, et al. Skin pigmentation related variants in Mexican population and interaction
670 effects on serum 25(OH)D concentration and vitamin D deficiency. Sci Rep. 29 juill
671 2024;14(1):17378.

672 23. Kerr R, Stevens G, Manga P, Salm S, John P, Haw T, et al. Identification of P gene mutations in
673 individuals with oculocutaneous albinism in sub-Saharan Africa. Hum Mutat. 2000;15(2):166-72.

674 24. Mauri L, Manfredini E, Del Longo A, Veniani E, Scarcello M, Terrana R, et al. Clinical evaluation
675 and molecular screening of a large consecutive series of albino patients. J Hum Genet. févr
676 2017;62(2):277-90.

677 25. Shahzad M, Yousaf S, Waryah YM, Gul H, Kausar T, Tariq N, et al. Molecular outcomes, clinical
678 consequences, and genetic diagnosis of Oculocutaneous Albinism in Pakistani population. Sci
679 Rep. 7 mars 2017;7:44185.

680 26. Sassone F, Estay-Ahumada C, Roux MJ, Ciocca D, Rossolillo P, Birling MC, et al. Interruption of the
681 visual cycle in a novel animal model induces progressive vision loss resembling Stargardts
682 Disease. Sci Rep. 28 déc 2024;14(1):30880.

683 27. Verra DM, Sajdak BS, Merriman DK, Hicks D. Diurnal rodents as pertinent animal models of
684 human retinal physiology and pathology. Prog Retin Eye Res. 1 janv 2020;74:100776.

685 28. Horowitz DS, Krainer AR. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA
686 splicing. Trends Genet TIG. mars 1994;10(3):100-6.

687 29. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational
688 constraint map using variation in 76,156 human genomes. Nature. 1 janv
689 2024;625(7993):92-100.

690 30. Andersen JD, Johansen P, Harder S, Christoffersen SR, Delgado MC, Henriksen ST, et al. Genetic
691 analyses of the human eye colours using a novel objective method for eye colour classification.
692 Forensic Sci Int Genet. sept 2013;7(5):508-15.

693 31. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific
694 haplotype structure and linking correlated alleles of possible functional variants. Bioinforma Oxf
695 Engl. 1 nov 2015;31(21):3555-7.

696 32. Mesdaghi S, Murphy DL, Simpkin AJ, Rigden DJ. Structural insights into pink-eyed dilution protein
697 (Oca2). Biosci Rep. 26 juill 2023;43(7):BSR20230060.

698 33. Pagani F. New type of disease causing mutations: the example of the composite exonic
699 regulatory elements of splicing in CFTR exon 12. Hum Mol Genet. 15 mai 2003;12(10):1111-20.

700 34. Pagani F, Raponi M, Baralle FE. Synonymous mutations in CFTR exon 12 affect splicing and are
701 not neutral in evolution. Proc Natl Acad Sci. 3 mai 2005;102(18):6368-72.

702 35. Raponi M, Baralle FE, Pagani F. Reduced splicing efficiency induced by synonymous substitutions
703 may generate a substrate for natural selection of new splicing isoforms: the case of CFTR exon
704 12. Nucleic Acids Res. 14 déc 2006;35(2):606-13.

705 36. Mueller WF, Larsen LSZ, Garibaldi A, Hatfield GW, Hertel KJ. The Silent Sway of Splicing by
706 Synonymous Substitutions. J Biol Chem. nov 2015;290(46):27700-11.

707 37. Kidd KK, Pakstis AJ, Donnelly MP, Bulbul O, Cherni L, Gurkan C, et al. The distinctive geographic
708 patterns of common pigmentation variants at the OCA2 gene. Sci Rep. 22 sept 2020;10(1):15433.

709 38. Currant H, Hysi P, Fitzgerald TW, Gharahkhani P, Bonnemaijer PWM, Senabouth A, et al. Genetic
710 variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence
711 tomography images. Hauser MA, éditeur. PLOS Genet. 12 mai 2021;17(5):e1009497.

712 39. Liu J, Bitsue HK, Yang Z. Skin colour: A window into human phenotypic evolution and
713 environmental adaptation. Mol Ecol. 2024;33(12):e17369.

714 40. Sauer DB, Song J, Wang B, Hilton JK, Karpowich NK, Mindell JA, et al. Structure and inhibition
715 mechanism of the human citrate transporter NaCT. Nature. mars 2021;591(7848):157-61.

716 41. Belcher CN, Vij N. Protein Processing and Inflammatory Signaling in Cystic Fibrosis: Challenges
717 and Therapeutic Strategies. Curr Mol Med. févr 2010;10(1):82-94.

718 42. Green DJ, Michaud V, Lasseaux E, Plaisant C, UK Biobank Eye and Vision Consortium, Fitzgerald T,
719 et al. The co-occurrence of genetic variants in the TYR and OCA2 genes confers susceptibility to
720 albinism. Nat Commun. 30 sept 2024;15(1):8436.

721 43. Michaud V, Lasseaux E, Green DJ, Gerrard DT, Plaisant C, UK Biobank Eye and Vision Consortium,
722 et al. The contribution of common regulatory and protein-coding TYR variants to the genetic
723 architecture of albinism. Nat Commun. 8 juill 2022;13(1):3939.

724 44. Sergouniotis PI, Michaud V, Lasseaux E, Campbell C, Plaisant C, Javerzat S, et al. A multilayered
725 approach to the analysis of genetic data from individuals with suspected albinism. J Med Genet.
726 déc 2023;60(12):1245-9.

727 45. Preising MN, Forster H, Gonser M, Lorenz B. Screening of TYR, OCA2, GPR143, and MC1R in
728 patients with congenital nystagmus, macular hypoplasia, and fundus hypopigmentation
729 indicating albinism. Mol Vis. 15 avr 2011;17:939-48.

730 46. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with
731 deep phenotyping and genomic data. Nature. 11 oct 2018;562(7726):203-9.

732 47. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making
733 protein folding accessible to all. Nat Methods. juin 2022;19(6):679-82.

734 48. Hekkelman ML, de Vries I, Joosten RP, Perrakis A. AlphaFill: enriching AlphaFold models with
735 ligands and cofactors. Nat Methods. févr 2023;20(2):205-13.

736 49. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction
737 with AlphaFold-Multimer [Internet]. bioRxiv; 2022 [cité 16 mai 2025]. p. 2021.10.04.463034.
738 Disponible sur: https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2

739 50. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein
740 structure prediction with AlphaFold. Nature. août 2021;596(7873):583-9.

741 51. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis
742 of massive data sets. Nat Biotechnol. nov 2017;35(11):1026-8.

743 52. Jaworek TJ, Kausar T, Bell SM, Tariq N, Maqsood MI, Sohail A, et al. Molecular genetic studies
744 and delineation of the oculocutaneous albinism phenotype in the Pakistani population. Orphanet
745 J Rare Dis. déc 2012;7(1):44.

746

747 # Figure captions

748

749 **Fig 1 Minigene assay of rare missense VUS in exon 10.**

750 (a) Position of variants that were tested by minigene assay. Variants in red have been shown
751 pathogenic due to excessive exon skipping in our previously reported study (12) whereas variants in
752 green had no significant effect. Missense variants tested in (c) are indicated in blue. The position of the
753 frequent benign variant c.1065G>A/p.A355= (rs1800404:C/T) is highlighted in yellow. Note that all rare
754 variants were tested in *cis* to c.1065G>A which is predominant in European populations and in our
755 patient cohort. (b) Minigene construct and expected sizes of the RT-PCR products with (335 bp) or
756 without (263 bp) exon 10. (c) Representative electrophoresis profile of RT-PCR products from variants
757 tested in minigene assay compared to control and pathogenic synonymous variant c.1080C>T;
758 p.Ser360=, using Agilent 4150 TapeStation instrument and D1000 DNA ScreenTape system. NTC: no-
759 template control. Relative quantification of exon 10 skipping in 4 series of independent experiments is
760 represented by histograms below the image. Mean % of skipping is indicated below each histogram.
761 All of the tested variants significantly increase exon 10 skipping except for c.1061T>C; p.Leu354Pro
762 which decreased skipping. Error bars represent SEM (n = 4). ***indicates p< 0.001, **** indicates p<
763 0.0001, one-way ANOVA followed by the Bonferroni post hoc multiple comparison test.

764

765 **Fig 2. Minigene assay profiling of human and mouse combined exon 10 sequences.**

766 (a) Alignment of human and mouse *OCA2* exon 10 (uppercase) and flanking intronic (lowercase)
767 sequences (human: ENSG00000104044; mouse: ENSMUSG00000030450) showing high exon sequence
768 homology (~82%). Sequences alignment was performed on Clustal Omega
769 (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Conserved nucleotides are indicated with an
770 asterisk. Amino acids and their position in the human protein are indicated above the alignment. The
771 two amino acid that are different in mouse are indicated below. Nucleotides indicated in blue have
772 been targeted for minigene constructs. The position of the frequent benign variant

773 c.1065G>A/p.A355= (rs1800404:C/T) is highlighted in yellow. (b) RT-PCR products using RNAs from
774 human melanocytes or mouse skin and primers in *OCA2* exon 9 and 13 (see Methods). The two bands
775 from the human sample correspond to the 2 transcripts with (383 bp) or without (311 bp) exon 10,
776 whereas in mouse all transcripts retain exon 10 resulting into a single RT-PCR band (424bp). NTC = no-
777 template control. (c) Human and mouse splicing profile by minigene assay. The two bands from the
778 human minigene correspond to the 2 transcripts with (335 bp) or without (263 bp) exon 10, whereas
779 with the mouse gene derived construct, all transcripts retain exon 10 resulting into a single RT-PCR
780 band (335 bp). (d) Summary of the effect of nucleotide swapping between human and mouse exon 10
781 sequences in the minigene assay. Minigene constructs with a human backbone are framed in purple.
782 Minigenes with a mouse backbone are framed in orange. The changes are highlighted using the
783 corresponding colours, experimental data in Fig S3.

784

785 **Fig 3. Minigene assay profiling of human and mouse intronic sequences.**

786 (a) Agarose gel electrophoresis of RT-PCR products from minigene assay on human, mouse and
787 chimeric minigene constructs. Human (dark purple) or mouse (dark orange) exon 10 sequences were
788 combined with only one or both human (light purple) or mouse (light orange) introns. (b) RT-PCR
789 products of minigene assay using a minimal mouse minigene with 75bp of intron 9 and 20 bp of intron
790 10. (c) Minigene assay assessing the effect of c.1045-9(G) (depicted with a red star), depending on the
791 surrounding mouse or human sequences. (d) Representative RT-PCR products profile from minigene
792 assay using human minigene with varying intron sizes. From left to right: (intron 9: 326bp/intron 10:
793 429bp); (73bp/429bp); (326bp/28bp); (73bp/28bp); (73bp/330bp); (73bp/212bp); (73bp/85bp) (upper
794 panel). Relative quantification of exon 10 skipping after transfection with the minigene constructs
795 above. Mean % of skipping is indicated below each histogram. Error bars represent SEM (n = 3). ****
796 indicates p<0.0001,** indicates p<0.01, ns indicates non-significant, one-way ANOVA followed by the
797 Bonferroni post hoc multiple comparison test (lower panel). NTC: no-template control.

798

799 **Fig 4. ASO mapping assay of human *OCA2* exon 10 combined with human or mouse intronic**
800 **sequences.**

801 (a) Location of the human *OCA2* exon 10 sequences targeted by each of the four 22-base long ASOs.
802 (b-e) for each condition, a representative screen tape electrophoresis is shown (upper) as well as the
803 relative quantification of exon 10 skipping obtained after co-transfection with each ASO, mean % of
804 skipping is indicated below each histogram (lower). Controls of the minigene assay are the result of
805 transfection with the minigene only. NT: non-targeting oligonucleotide control, V: transfection vehicle
806 control, NTC: no-template control. One-way ANOVA followed by the Bonferroni post hoc multiple
807 comparison test, **** indicates p<0.0001, ns indicates non-significant. Error bars represent SEM (n =
808 3). For multiple comparison, same letters indicate no significant differences, na: non-applicable (n=2).
809 (b-c) human minigene with 50nM (b) or 5nM (c) ASOs. (d-e) chimeric minigene with mouse introns and
810 human exon with 50nM (d) or 5nM (e) ASOs.

811

812 **Fig 5. Functional analysis of the effect of c.1065G>A on exon 10 skipping by minigene assay.**

813 Representative screen tape electrophoresis of RT-PCR products from compared minigene assays of
814 constructs with c.1065G or c.1065A. The additional effect of c.1080C>T/p.Ser360= is tested in *cis* to
815 each of these variants. V: transfection vehicle control. NTC: no-template control. Relative
816 quantification of exon skipping showing significant increase in skipping with c.1065A compared to
817 c.1065G and the additional effect of c.1080C>T. Error bars represent SEM (n = 4). A one-way ANOVA
818 followed by the Bonferroni post hoc multiple comparison test was performed. Variables with different
819 letters are significantly different (p<0.0001) (lower panel).

820

821 **Fig 6. Haplotypes derived from rs1800404:C/T and rs1800407:C/T to skin and hair colour in European**
822 **subjects.**

823 (a-b) Self-reported categorical skin (a) and hair (b) colour distribution in the studied UK Biobank
824 population (n= 402871). (c) Distribution of the 9 possible genotypes for rs1800404:C/T and
825 rs1800407:C/T in the studied population. Only 2 out of 402871 subjects are double homozygotes for
826 the variants of each SNVs ($f=5.10^{-6}$). (d-e) Effects of rs1800404 and rs1800407 haplotypic combinations
827 expressed as odd ratios. Each dot corresponds to the effects of *OCA2* haplotypes on skin/hair
828 pigmentation considering the rs1800404:C/rs1800407:C haplotype as a reference. Major categories in
829 the studied population (fair for skin, and light brown for hair colour) are used as phenotypical
830 references (red dashed vertical line). Haplotype effects were obtained from a multinomial regression
831 model adjusted for age and sex. (d) all subjects were included, n=402871, 6 categories. (e) Subjects
832 with red hair or non-reported hair colour were excluded, n= 379183, 4 categories.

833

834 **Fig 7. Conformation prediction of the ΔE10-OCA2 protein compared to the native OCA2 protein.**

835 (a) 2D transmembrane topology (upper panel) of the OCA2 protein (left) and the ΔE10-OCA2 protein
836 (right) and corresponding structural models (lower panel) of monomers (ColabFold). The OCA2 protein
837 topology is shown as transparent on the ΔE10-OCA2 topology. Boxes highlight the pseudo inverse
838 repeat with scaffold and transport domains; Blue/black indicate alternating coding exons.
839 Glycosylation sites correspond to N214, N218, N273; Dileucine motives are at position p.96-102, p.135-
840 140, p.154-159. Structural models of the OCA2 and ΔE10-OCA2 monomers highlighting exon 9 (blue),
841 exon 10 (purple), and exon 11 (yellow) encoded domains. The cytosolic N-terminal 165 residue stretch
842 with undefined conformation is not included in the structural models. (b) Structural models of three
843 AlphaPulldown dimer predictions starting at residue 166. Each model is shown in cartoon
844 representation with both chains visible. Upper panel: Models are coloured according to the pLDDT
845 scale, where dark blue indicates high confidence (>90), light blue represents moderate confidence (70–

846 90), yellow low confidence (50–70), and orange very low confidence (<50). Lower panel: same models
847 highlighting exon encoded domains as in (a).

848

849 **Fig 8. *OCA2* exon 10 skipping and pigmentation across a continuous spectrum from pathology to**
850 **physiology.**

851 Rare variants such as c.1045-15T>G (red star) result in full skipping of *OCA2* exon 10 whether in *cis* to
852 c.1065G (52) or c.1065G>A (12). Rare variants such as c.1080C>T that enhance skipping levels leaving
853 some transcripts that retain exon 10 (orange star) are pathogenic when in *cis* to c.1065G>A
854 (rs1800404-T). This frequent variant increases basal levels of exon 10 skipping on its own and is
855 benignly associated to lightening of the skin and hair in healthy individuals of all ethnicities including
856 with European ancestries (this study). The effect of rs1800404-T is dose-dependent as homozygotes
857 have a significantly lighter skin and hair than heterozygotes. The retinal pigmentation is likely affected
858 given the contribution of rs1800404-T (c.1065G>A) to the ocular pathology in patients with OCA.
859 Patients with only one identified pathogenic variant should be investigated for haplotypic
860 combinations that would increase skipping in the transcripts encoded by the second allele (blue box).
861 The probability of such a pathogenic haplotype will depend on the patient's ethnicity (for instance
862 c.1114G>A/Asp372Asn is significantly more frequent in South Asia). A *HERC2-OCA2* "lightening"
863 haplotype (in favour of increased exon 10 skipping ratio) is expected to contribute to modulation of
864 both physiological pigmentation and severity of OCA2 depending on the second allele. Given the
865 extreme sensitivity of the system to skipping levels, it can be expected that such a haplotype will
866 influence the severity of other OCAs such as OCA1 driven by hypomorphic variants of *TYR*.

867

868

869 # Supporting information

870 **Fig S1. Robust quantification of skipping levels by minigene assay using DNA Screentape analysis.**

871 (a), Whole gel image of RT-PCR products corresponding to Fig 1b using D1000 DNA Screentape system
872 on an Agilent TapeStation 4150 instrument. PCR products derived from minigene transcripts are
873 detected through fluorescence around 300 bp. The upper (1500 bp) and lower (25 bp) markers, in
874 D1000 loading dye, are size markers used as references for assessing the DNA fragments sizes in a
875 sample. NTC= no-template control. (b) Electropherogram obtained following electrophoresis of RT-
876 PCR products from the control minigene, shown in the gel above. The percentage integrated area (PIA)
877 of the peak corresponding to the RT-PCR product without exon 10 is directly used to estimate the % of
878 exon skipping as illustrated in the table below.

879

880 **Fig S2. Alignment of *OCA2* sequences in and around exon 10 (or its equivalent) from different species**
881 **that do or do not express the transcript without exon 10 (or its equivalent).**

882 (a) Alignment of *OCA2* intron 9 sequences. Sequences alignment was performed on Clustal Omega
883 (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Conserved nucleotides are identified by * while
884 non homologous nucleotides are indicated by -. The human sequence is hightlighted in yellow and the
885 putative splicing branch point is in red 47 bp upstream the intron 3' splice site (note that for mouse,
886 the branch point is predicted at -37 bp, not shown on the alignment). Intronic nucleotides near the
887 exon highlighted in blue indicate highly conserved sequences across mammalian species. (b) Alignment
888 of *OCA2* exon 10 sequences. Codon positions are indicated above the alignment. Non conserved
889 nucleotides are highlighted: green for T, pink for C, blue for G, red for A. (c), Alignment *OCA2* intron
890 10, legend as for (a).

891

892 **Fig S3. Experimental results of minigene assays resulting from nucleotide swapping between human**
893 **and mouse exon 10 sequences supporting Fig 2d.**

894 (a) Products of RT-PCR performed on mouse skin or optic cup RNAs using primers in exon 9 and 13 (see
895 Methods) (b-c) Agarose gel electrophoresis of RT-PCR products of human (b) and mouse (c) minigene
896 assay showing the effect of nucleotide swapping between human and mouse exon 10 sequences.
897 Sequences of human origin are indicated in purple; sequences of murine origin, in red. NTC= no-
898 template control.

899

900 **Fig S4. Transcript profile of *OCA2* in *Psammomys obesus* (Fat Sand Rat).**

901 (a) Alignment between *OCA2* exon 10 and surrounding sequences of human, mouse and *Psammomys*
902 *obesus* (human: ENSG00000104044; mouse: ENSMUSG00000030450; *Psammomys*: NW 026590486.1
903 (1302928-1626514)). Sequence alignment was performed on Clustal Omega
904 (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Upper case letters refer to exon sequence (also
905 highlighted in yellow) and lower case letters to intronic sequences flanking the exon. Intronic
906 nucleotides highlighted in blue indicate similar sequences between the species, red indicates
907 homology between mouse and *Psammomys* but not human whereas pink indicates nucleotides that
908 are common to human and one of the 2 rodent species. (b), Agarose gel electrophoresis of RT-PCR
909 products of *Oca2* from P*sammomys* skin and eye total RNAs (primers in exon 8 and 12, see Table S1).
910 RT-PCR products correspond to the transcript correctly spliced with all exons (414bp). This experiment
911 was done on the skin of three different animals (1 to 3) and on one eye of animal n°1. NTC = no-
912 template control.

913

914 **Fig S5. Cumulative effect of synonymous c.1047C>T/p.Ile349=.**

915 Minigene assay evaluated by conventional agarose electrophoresis showing the additional effect of
916 c.1047C>T/p.Ile349= equivalent to the increase in skipping induced by c.1080C>T that was precisely
917 quantified by Screen tape analysis (Fig 5). C.1047C>T is comparatively tested in *cis* to c.1065G or
918 c.1065A as indicated. NTC: no-template control.

919

920 **Fig S6. Skin or hair colour categories association to genotypes at rs1800407 and rs1800404 in the UK**
921 **Biobank population with European ancestry.**

922 (a-b) Comparative association of rs1800407:C/T and rs1800404:C/T to skin colour. (c-d) Comparative
923 association of rs1800407:C/T and rs1800404:C/T to hair colour. Association of skin colour (a) or hair
924 colour (c) to the 3 genotypes (C/C, C/T, T/T) at rs1800407 (left) or rs1800404 (right) in the UK Biobank
925 with European ancestry (n=402871). Categories are indicated by corresponding colours. For each SNV
926 analyzed separately, significant association is detected between the (T) allele and paleness of the skin
927 (a). However, for the hair colour (c), the effect of rs1800407:C/T is opposite to what is expected, with
928 (T) associated to darker rather than lighter hair colour. The effect of rs1800404-T, on the other hand,
929 is in agreement with a contribution to paleness of both skin and hair (see text for statistical values). (b-
930 d) Association of skin colour (b) or hair colour (d) to the 9 combinations of genotypes at rs1800404:C/T
931 and rs1800407:C/T. Double heterozygotes with *cis* or *trans* segregation of the variants cannot be
932 discriminated. This illustrates the strong association of rs1800404-T with hypopigmentation of the skin
933 and hair independently of rs1800407-T as statistically shown in Fig 6.

934

935 **Fig S7. Compared statistical analysis for conformation and dimerization prediction of the native**
936 **OCA2 protein and the ΔE10-OCA2 protein.**

937 (a) Structural models of two ColabFold OCA2 monomers coloured by per-residue confidence (pLDDT).
938 Models are truncated removing the first 165 residues representing the N-terminal disordered region.

939 Models are coloured according to the pLDDT scale, where dark blue indicates high confidence (>90),
940 light blue represents moderate confidence (70–90), yellow low confidence (50–70), and orange very
941 low confidence (<50). Each model is shown in cartoon representation with both chains visible. Below:
942 Predicted Aligned Error (PAE) maps the two monomer models. PAE maps visualise the predicted
943 alignment error (in Angstroms) between residue pairs across and within chains. Blue regions indicate
944 high confidence (low error), and red regions represent low confidence (high error), with values capped
945 at 30 Å. Diagonal blocks correspond to intrachain predictions; off-diagonal blocks indicate interchain
946 confidence. (b) PAE maps for three AlphaPulldown dimeric models. Same as above. Below:
947 Comparative summary of PAE maps for the three dimer AlphaPulldown models. Each model is
948 evaluated for intrachain confidence and interchain interaction quality based on PAE colour
949 distribution. Intrachain confidence reflects the predicted accuracy of the individual monomer folds,
950 while interchain confidence and interface clarity indicate the reliability and extent of the predicted
951 dimer interface.

952

953 **Fig S8. Haplotypic combinations involving a rare missense variant of exon 10 and c.1065G>A;**
954 **rs1800404:C/T in GnomAd.**

955 Sequencing data (GnomAd V4.0.1) from subjects carrying a rare missense or synonymous variant of
956 exon 10 was examined for segregation with c.1065G>A. p.V350M (2 homozygotes); p.L354P (no
957 homozygote); p.A355V (4 homozygotes); p.L361V (no homozygote); p.A368V (1 homozygote);p. I370T
958 (9 homozygotes); p.G371= (479 homozygotes); p.D372N (1 homozygote). On all recovered sequences
959 (n=156), half were incomplete such that the haplotype could not be defined. On the remaining: all
960 presumably healthy homozygotes carried the rare variant segregating in *cis* to c.1065(G), n=6: 2 with
961 p.V350M and 4 with p.A355V (see corresponding sequence panels). As for the heterozygous carriers,
962 only 3 heterozygotes for the same variant (D372N) carried the variant in *cis* to c.1065G>A (last panel).
963 Two were double heterozygotes and one was homozygote for c.1065G>A. All in all, no presumably

964 healthy homozygote for a rare missense variant that enhances exon 10 skipping carried the variant in
965 cis to c.1065G>A in agreement with the benign nature of such haplotype. The position of c.1065 is
966 indicated by a blue arrow, the position of the rare variant by a red arrow.

967

968 **Table S1. Oligonucleotidic sequences of all primers used for mutagenesis, cloning or RT-PCR**
969 **reactions.**

970 **Table S2. OCA2 transcript identification for the species that were included in this study.**

971 Note: Informations were collected on Ensembl: http://www.ensembl.org/index.html except for
972 *Psammomys obesus* informations were obtained on NCBI: https://www.ncbi.nlm.nih.gov/. Abbreviations:
973 Δ10=exon 10 skipped.

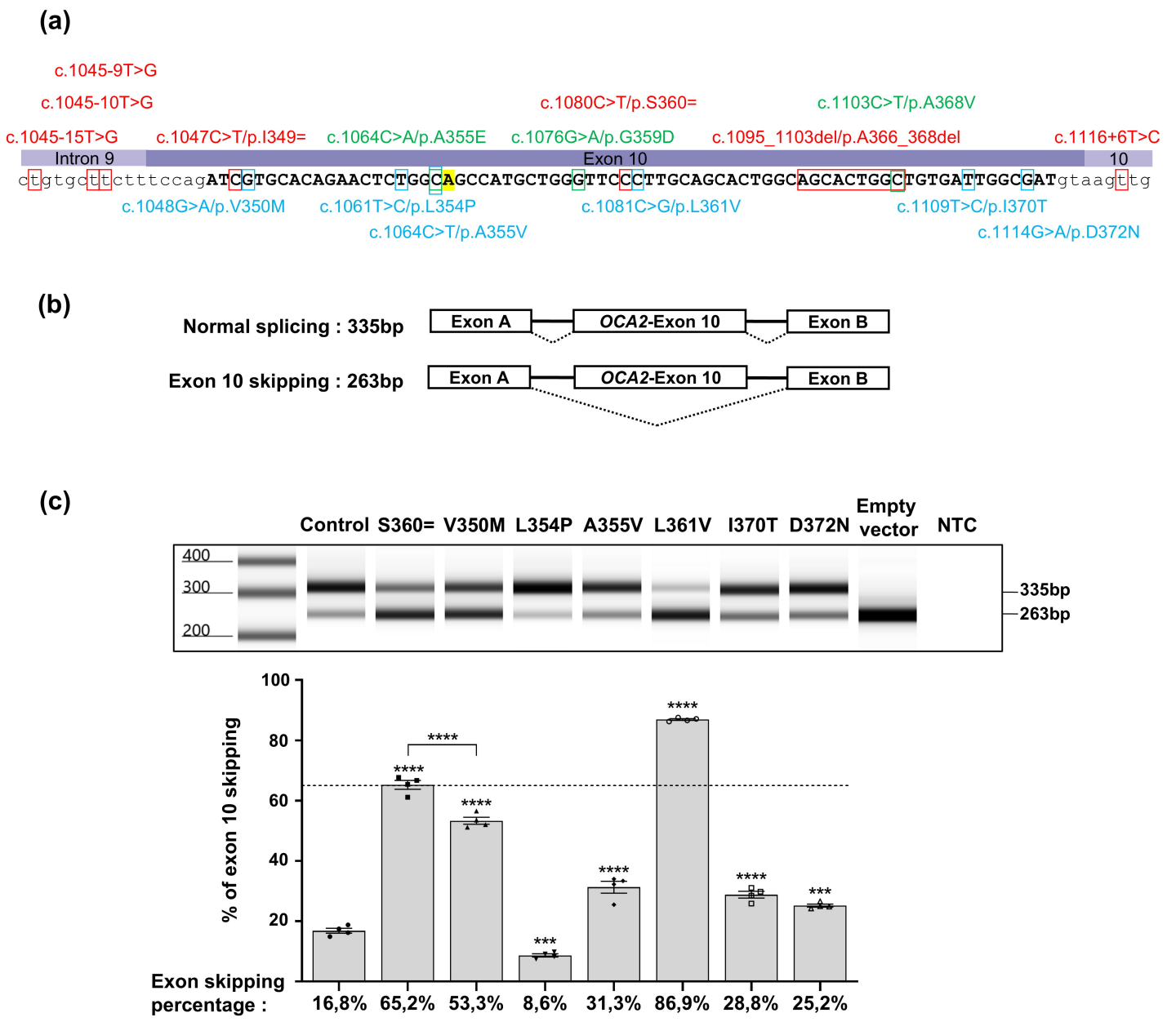
974 **Table S3. Statistical data for rs1800404 and rs1800407 haplotypic association studies performed on**
975 **the UK Biobank population with European ancestry.**

976 (a) association to skin colour. (b) association to hair colour.

977 **Table S4. Allelic frequencies of selected SNVs of exon 10 illustrating high degree of variability across**
978 **the geographical subpopulations.**

979 For each variant, the population with the highest frequency is highlighted in red, the one with the
980 lowest frequency in green. NA: non applicable (too many citations).

981

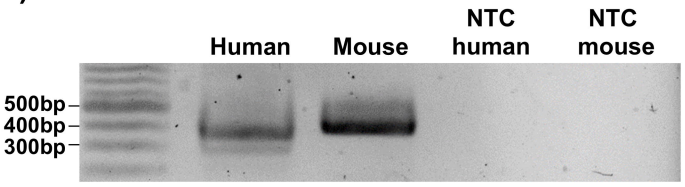

(a)
c.1045-9T>G
c.1045-10T>G
c.1080C>T/p.S360=
c.1103C>T/p.A368V
c.1045-15T>G
c.1047C>T/p.I349=
c.1064C>A/p.A355E
c.1076G>A/p.G359D
c.1095_1103del/p.A366_368del
c.1116+6T>C
Intron 9
Exon 10
10
ctgtgcttctttccagATCGTGCACAGAACTCTGGCAGCCATGCTGGGTTCCCTTGCAGCACTGGCAGCACTGGCTGTGATTGGCGATgtaagttg
c.1048G>A/p.V350M
c.1061T>C/p.L354P
c.1081C>G/p.L361V
c.1109T>C/p.I370T
c.1064C>T/p.A355V
c.1114G>A/p.D372N
(b)
Normal splicing : 335bp
Exon A
OCA2-Exon 10
Exon B
Exon 10 skipping : 263bp
Exon A
OCA2-Exon 10
Exon B
(c)
Control
S360=
V350M
L354P
A355V
L361V
I370T
D372N
Empty vector
NTC
400
300
200
335bp
263bp
% of exon 10 skipping
100
80
60
40
20
****
***
Exon skipping percentage :
16,8%
65,2%
53,3%
8,6%
31,3%
86,9%
28,8%
25,2%

(a)

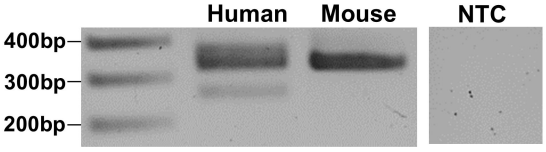
```
                                349        351        353        355        357        359
                                Ile  Val  His  Arg  Thr  Leu  Ala  Ala  Met  Leu  Gly  Ser
Human   cttctttccagATC GTG CAC AGA ACT CTG GCA GCC ATG CTG GGT TCC
Mouse   ttcctttccagATT GTT CAC AGA ACC CTG GCA GCC ATG TTG GGA GCA
        _*_**********_ **_ *** *** **_ *** *** *** *** _** **_ _*_
                                                                Ala

        361        363        365        367        369        371
        Leu  Ala  Ala  Leu  Ala  Ala  Leu  Ala  Val  Ile  Gly  Asp
Human   CTT GCA GCA CTG GCA GCA CTG GCT GTG ATT GGC GATgtaagttgtcac
Mouse   CTT GCA GCA CTA GCA GCC TTG GCT GTG GTT GGA GATgtaagttgttat
        *** *** *** **_ *** **_ _** *** *** _** **_ ************_*_
                                                Val
```

(b) RT-PCR

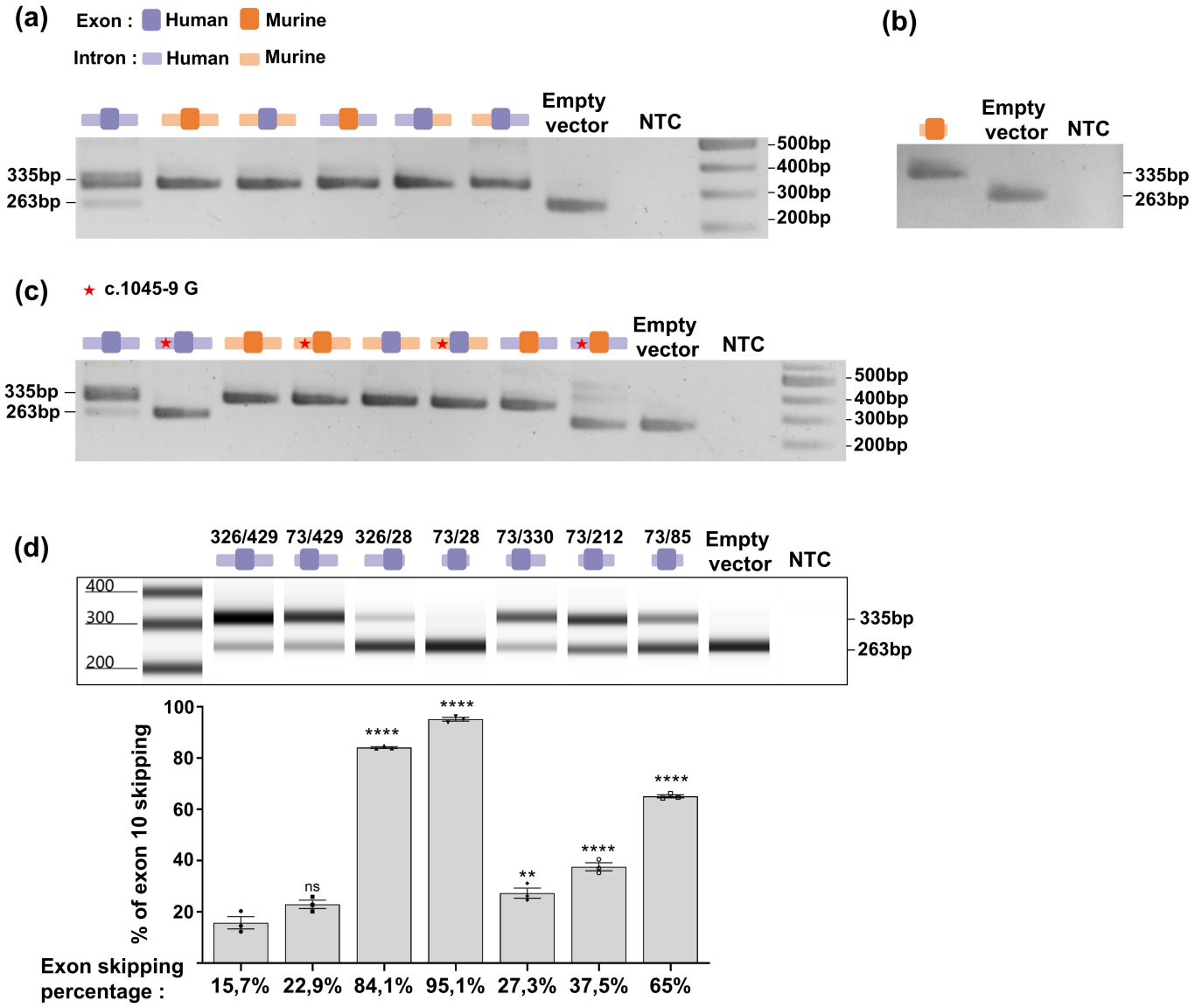


(c) Minigene assay



(d)

	Human	Skip	Mouse	Skip
p.349 WT	Ile (ATC)	15-20%	Ile (ATT)	No skip
	Ile (ATT)	**>40%**	Ile (ATC)	No skip
p.360 WT	Ser (TCC)	15-20%	Ala (GCA)	No skip
	Ser (TCA)	**>30%**	Ser (TCA)	No skip
	Ala (GCC)	**>70%**	Ser (TCC)	No skip
	Ala (GCA)	**>80%**		
p.370 WT	Ile (ATT)	15-20%	Val (GTT)	No skip
	Val (GTT)	15-20%	Ile (ATT)	No skip
p.360+p.370 WT	TCC+ATT	15-20%	GCA+GTT	No skip
			TCC+ATT	No skip


(a)
Exon : Human Murine
Intron : Human Murine
Empty vector
NTC
335bp
263bp
500bp
400bp
300bp
200bp
(b)
Empty vector
NTC
335bp
263bp
(c)
c.1045-9 G
Empty vector
NTC
335bp
263bp
500bp
400bp
300bp
200bp
(d)
326/429
73/429
326/28
73/28
73/330
73/212
73/85
Empty vector
NTC
400
300
200
335bp
263bp
% of exon 10 skipping
100
80
60
40
20
ns
****
****
**
****
****
Exon skipping percentage :
15,7%
22,9%
84,1%
95,1%
27,3%
37,5%
65%

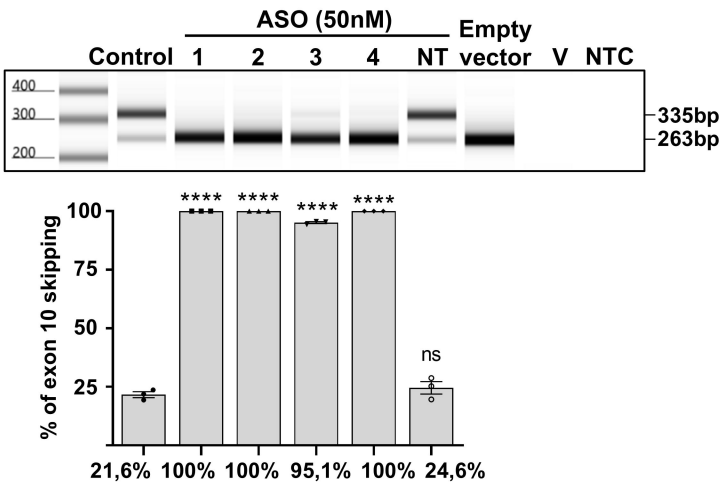
(a) ***OCA2* exon 10 sequence :**

ASO2 ASO4

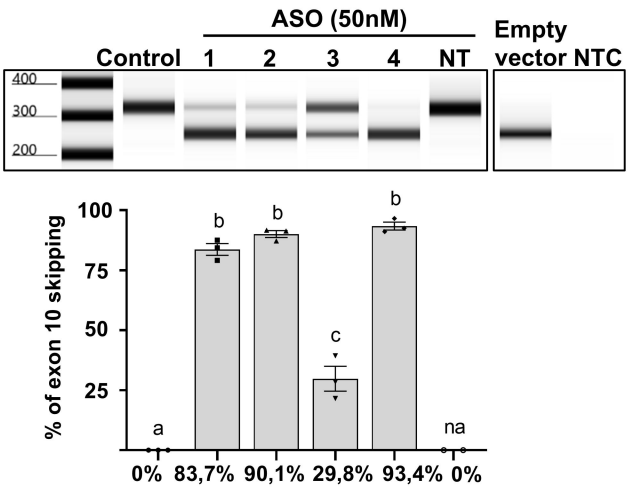
ATCGTGCACAGAACTCTGGCAGCCATGCTGGGTTCCCTTGCAGCACTGGCAGCACTGGCTGTGATTGGCGAT

ASO1 ASO3

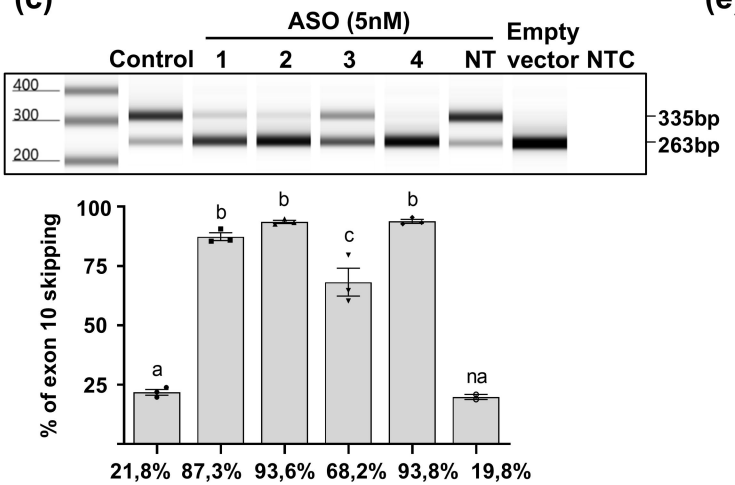
(b) Minigene construct :



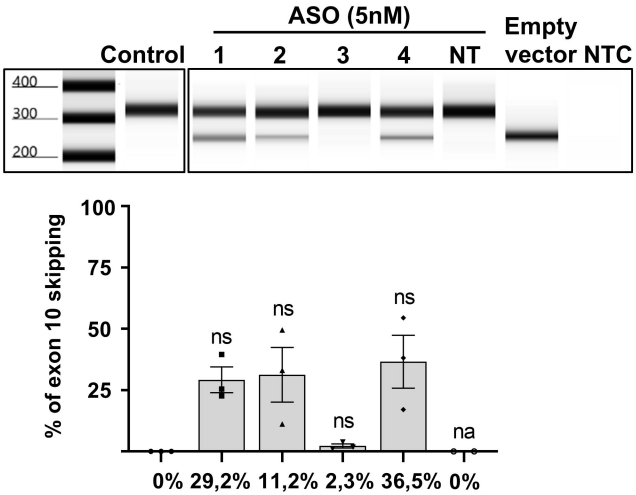
(d) Minigene construct :

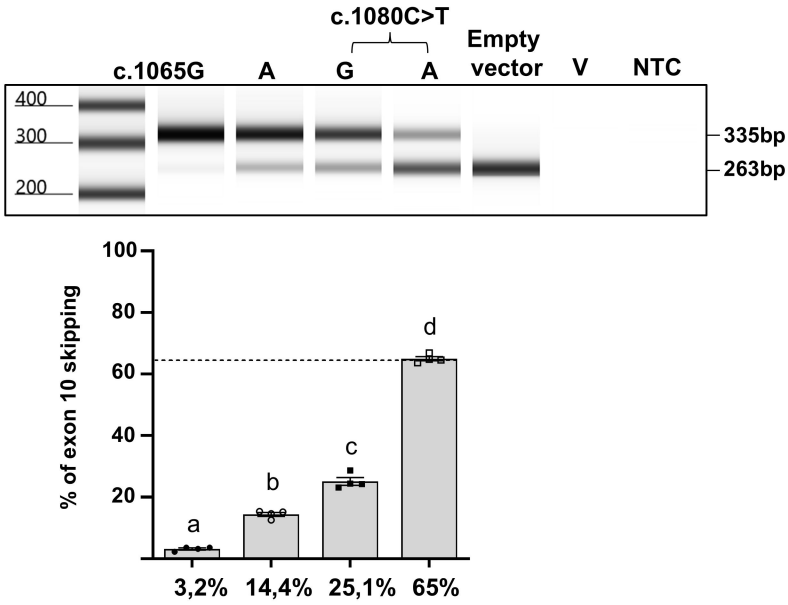


(c)

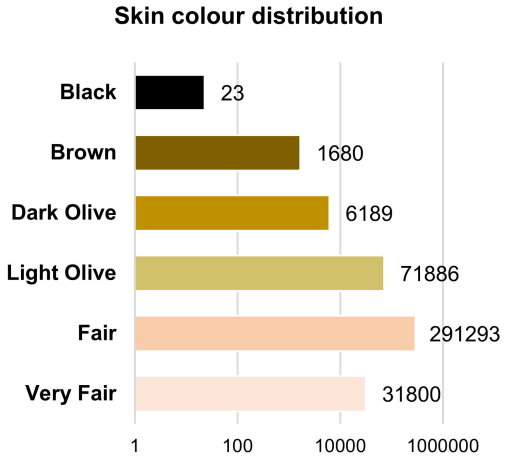


(e)

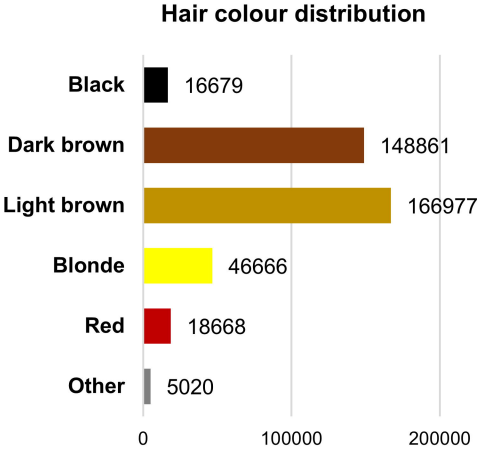



c.1080C>T
c.1065G
A
G
A
Empty vector
V
NTC
400
300
200
335bp
263bp
100
80
60
40
20
% of exon 10 skipping
a
b
c
d
3,2%
14,4%
25,1%
65%

(a)



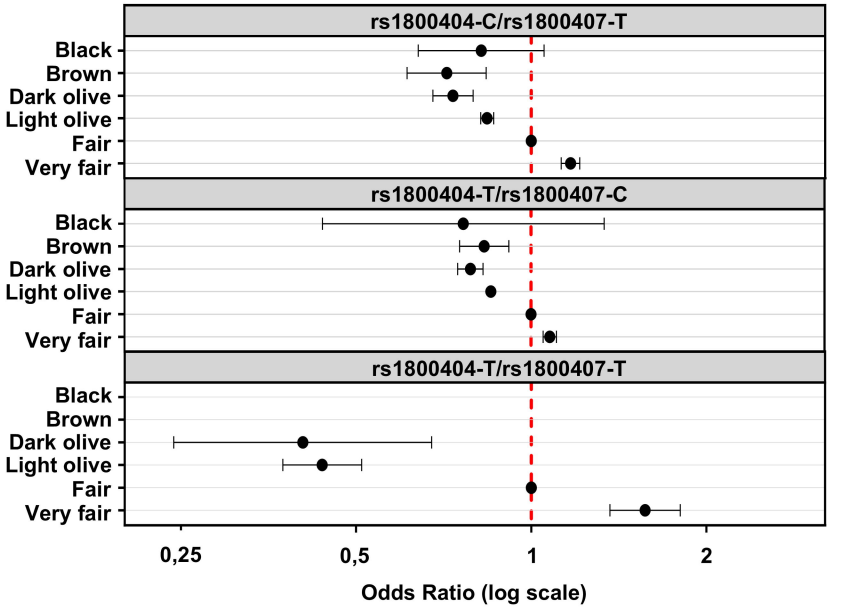
(b)



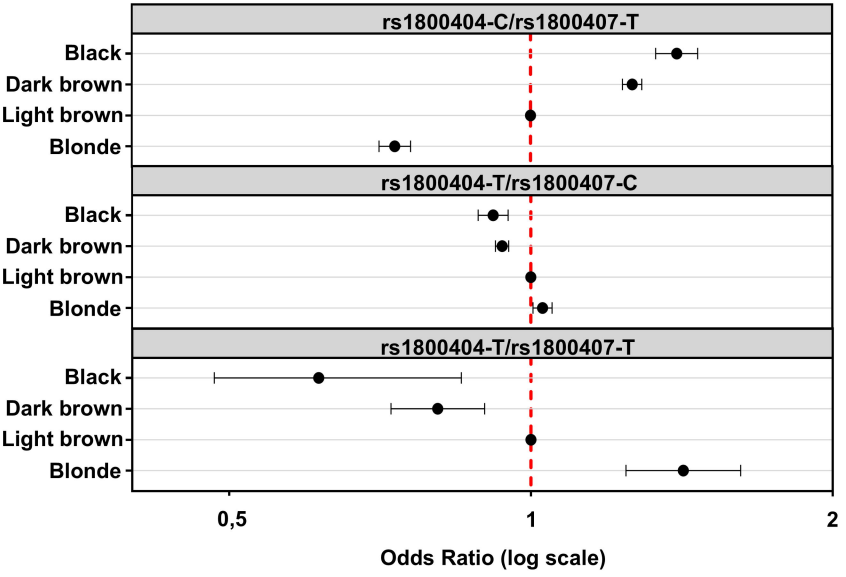
(c)

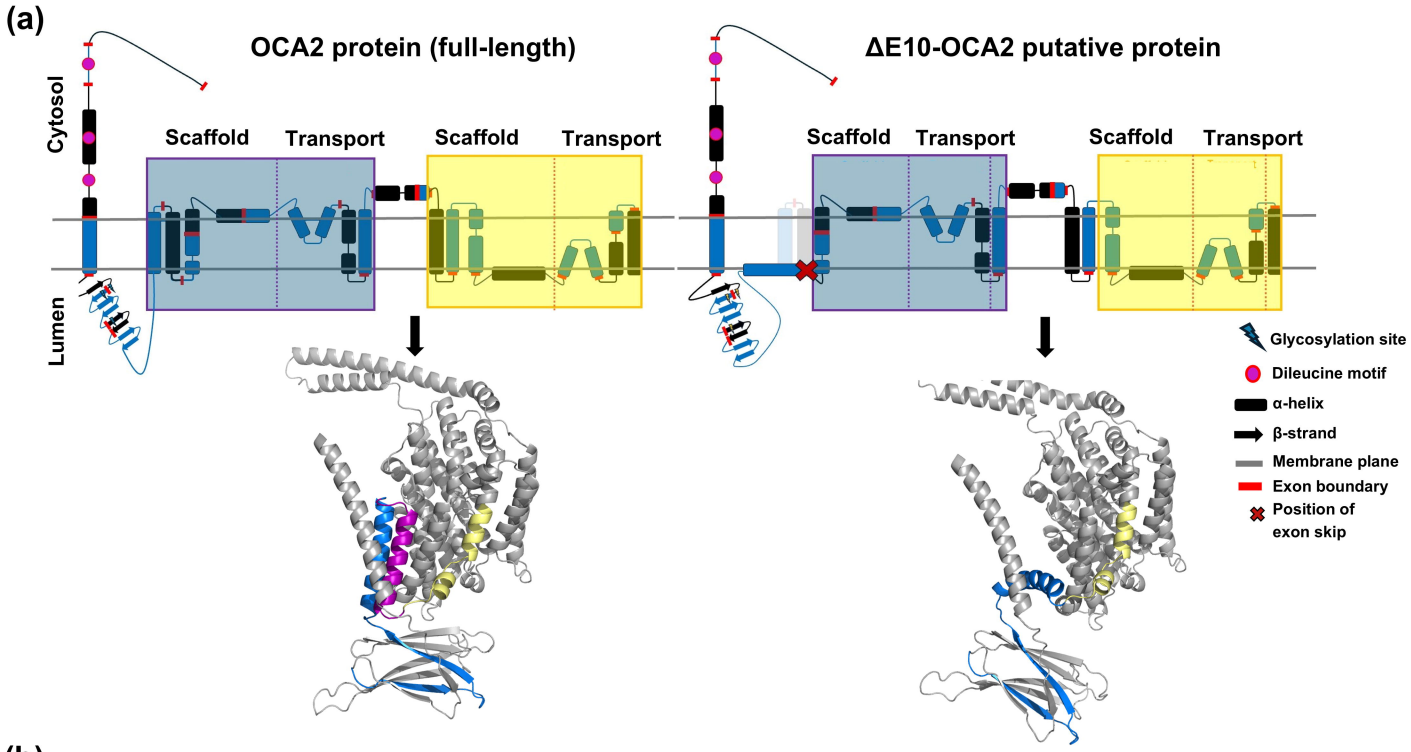
		rs1800404			
		C/C	C/T	T/T	total
rs1800407	C/C	6274	79544	251028	336846
	C/T	8276	53023	1698	62997
	T/T	2834	192	2	3028
	total	17384	132759	252728	402871

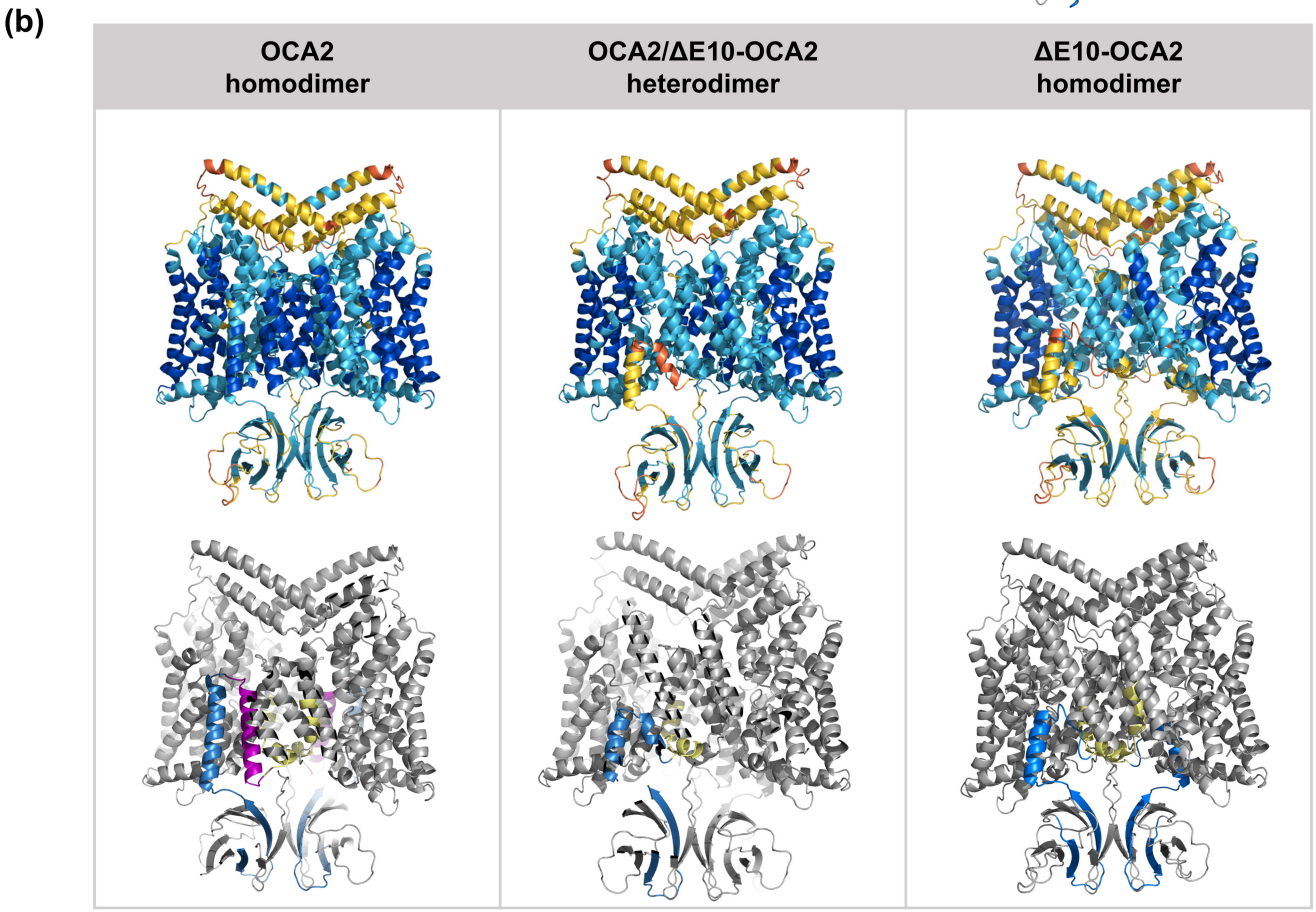
(d)

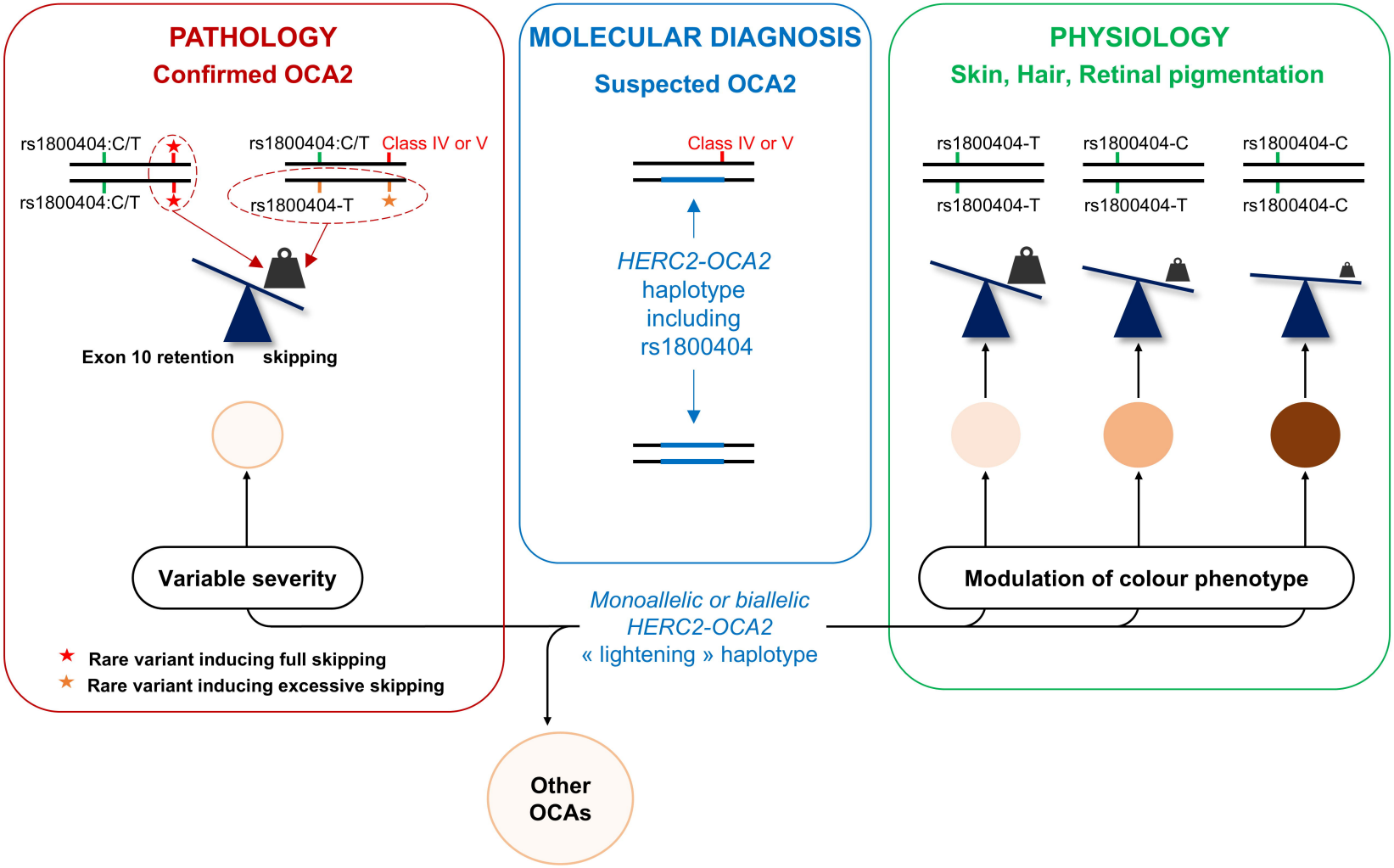


(e)




(a)
OCA2 protein (full-length)
ΔE10-OCA2 putative protein
Cytosol
Lumen
Scaffold
Transport
Scaffold
Transport
Scaffold
Transport
Scaffold
Transport
Glycosylation site
Dileucine motif
α-helix
β-strand
Membrane plane
Exon boundary
Position of exon skip


(b)
OCA2 homodimer
OCA2/ΔE10-OCA2 heterodimer
ΔE10-OCA2 homodimer


PATHOLOGY
Confirmed OCA2
rs1800404:C/T
rs1800404:C/T
rs1800404:C/T
Class IV or V
rs1800404-T
Exon 10 retention
skipping
Variable severity
Rare variant inducing full skipping
Rare variant inducing excessive skipping
MOLECULAR DIAGNOSIS
Suspected OCA2
Class IV or V
HERC2-OCA2
haplotype
including
rs1800404
PHYSIOLOGY
Skin, Hair, Retinal pigmentation
rs1800404-T
rs1800404-T
rs1800404-C
rs1800404-T
rs1800404-C
rs1800404-C
Modulation of colour phenotype
Monoallelic or biallelic
HERC2-OCA2
« lightening » haplotype
Other
OCAs